APPLIED MATHEMATICAL STATISTICS

By

R. L. Anderson
Institute of Statistics, N. C. State College

T. A. Bancroft
Statistical Laboratory, Iowa State College

September, 1949

Replaces
Institute of Statistics
Kimee Series #2
Errata

Chapter II

Page 9, line 9, chance was to were and seventeen to seventeenth.
Page 10, line 7b, chance implied to implies.
Page 12, line 2b, should read, "A bag contains 4 red ..."
Page 14, line 2 of last 5 should have comma after P.
Page 16, Exercise 2.9, add taking some or all to last sentence.
Page 16, Exercise 2.13, change $P_{50}^5 \frac{r-x}{5} \frac{6}{54}$ to $P_{50}^5 \frac{56}{54} \frac{r+6}{54}$
Page 17, Exercise 2.17, change the probabilities of A or B not happening are 3/11 and 2/7 respectively to the probability of A not happening is 3/11 and the probability of B not happening is 5/7.
Page 17, Exercise 2.21, change even to event.

Chapter III

Page 19, comma after normal in line 6b.
Page 19, Table 3.1 replace large X by small x.
Page 19, line 4b, change Fisher (1) to Fisher (3).
Page 23, second formula, for $g(x)$ omit one $x$ in denominator.
Page 23, line 7b, change $\frac{N^x}{x!} \frac{m}{x!} \text{ to } \frac{N^x}{x!} \frac{m}{x!}$
Page 24, line 9, change Snedecor (2) to Snedecor (7).
Page 25, Figure 3.5, move origin over to correspond to Figure 3.6.
Page 26, equation (3) should have integral sign.
Page 27, lines 3 and 7 omit of course.
Page 27, equation (6) and following, change u to $u$.
Page 27, line 8b, change $x + \sigma$ to $x = \mu + \sigma$.
Page 30, Exercise 3.1, change $\mu \sigma$ to $\mu \sigma$.
Page 30, Exercise 3.5, change $f(x)$ to $f(x)$.
Page 32, Omit Exercise 3.11.
Page 32, Exercise 3.16, change $f(x) = 2a(1 - x/a)$ to $f(x) = \frac{2a - x}{b^2}$
Page 33, omit Exercise 3.19.
Page 33, Exercise 3.21, change $\mu \sigma$ to $\mu \sigma$.
Page 34, Exercise 3.29, change $0 < x < 1$ to $0 < x < \infty$.
Page 34, omit Exercise 3.31.
Page 36, last line should have comma after $\phi(x)$.

Page 38, line 3b, change $\int (x - \mu) f(x) \, dx$ to $\int (x - \mu)^k f(x) \, dx$.

Page 39, line 13, write $\frac{\mu^k}{\sigma^k}$.

Page 40, line 5, change $k = \frac{c}{W}$ to $k = \sqrt{\frac{c}{W}}$.

Page 40, line 6, change $(x - \mu)^2$ to $(x - \mu)^2$.

Page 41, line 7, change $\theta(t)$ to $\theta'(t)$.

Page 41, line 5b, change $\mu_1$ to $\mu$.

Page 41, last line, write $E \left[ e^{t \theta(x)} \right]$.

Page 42, line 3, change $(1 - t)$ to $(1 - t)^{-1}$.

Page 42, line 5, change $u_k$ to $u_k^*$.

Page 42, line 11, change $\mu_1^{\mu} = \mu_1$ to $\mu = \mu$ etc.

Page 42, line 3b, change $E(e^{t(x - \mu)})$ to $E \left[ e^{t(x - \mu)} \right]$.

Page 42, line 2b, change $e^{-t \mu} E(e^{tW})$ to $e^{-t \mu} E(e^{tX})$.

Page 42, last line, change $e^{-t \mu} \phi(t)$ to $e^{-t \mu} \phi'(t)$.

Page 43, line 10b, change $\mu$ to $\mu_1$.

Page 43, line 6b, change $tu_1$ to $tu$.

Page 43, line 2b, change $\mu$ to $t\mu$, change last parenthesis

$\frac{1}{2}(\text{etc.}) = \frac{1}{2}$

Page 43, last line, change $\mu^{\mu}$ to $\mu$.

Page 44, Example 4.9, change $\frac{\mu^2}{2}$ to $\frac{1}{2 \sigma^2}$; change $e^{-tx}$ to $e^{tx}$.

Page 44, change $\phi(t)$ to $\phi'(t)$.

Page 46, Theorem 4.1, change exist to exists.

Page 48, Exercise 4.16, last line, change $x_1$ to $x_2$.

Page 49, Exercise 4.23 and 4.25, change $\mu_1$ to $\mu$. 
Errata
Chapter V

Page 52, Table 5.2, a vertical line is missing.
Page 54, line 4, change \( f(x) \) to \( q(x) \).
Page 54, line 7, change \( (x,y+dx) \) to \( (x,x+dx) \).
Page 54, equation at bottom of page, change \( 2pxy \) to \( 2p_{xy} \).
Page 56, line 12, omit \( q \) and change \( f_2(w) \) to \( f(w) \). Also drop subscript 2 in following table of values.
Page 58, Example 5.6, change \( v = xy \) to \( v = x/y \). Reverse the rows and columns of \( J \). Change \( \frac{(xy)}{y^2} \) to \( \frac{-xy}{y^2} \).
Page 60, change Example 5.6 to Example 5.7, change Example 5.7 to Example 5.8.
Page 61, change Example 5.6 to Example 5.9. Change Example 5.8 to Example 5.10.
Page 61, new Example 5.10, change \( \phi \) to \( \phi \). Change \( cy^2 \) to \( c^2y \).
Page 62, change Example 5.10 to Example 5.12.
Page 62, new Example 5.12, line 8b, second summation in \( \mu_{20} \) should be from 0 to 1 instead of from 0 to 2.
Page 63, for the cumulant-generating function change the capital \( K \) end small \( k \) to capital and small kappas respectively.
Page 64, equal mark after \( 1/J \).
Page 65, line 5, change \( \sum (p_{ij} e^{\theta_i}) \) to \( \sum (p_{ij} e^{\theta_i}) \).
Page 65, line 1b, omit \(-\lambda_u \).
Page 66, Exercise 5.1, last line, change \( y(x) \) to \( g(x) \).
Page 66, change Exercise 5.2 to Exercise 5.3 and vice versa.
Page 66, new Exercise 5.3, (b) rewrite as \( g(x) = \frac{2(a-x)}{x} \). (e) rewrite \( \mu_{01} = \frac{2a}{x} \), rewrite \( \mu_{y,x} = \frac{a+x}{2} \).
Page 66-67, Exercise 5.4, change "Exercise 5.3" to Exercise 5.2. Change \( y(x) \) to \( g(x) \). Last line change \( cy \) and \( cx \) to \( c_y \) and \( c_x \) respectively.
Page 67, Exercise 5.5, the \( f \) should be in numerator as \( f(x,y) \).
Page 68, Exercise 5.9b, change \( y(x) \) to \( g(x) \).
Page 69, Exercise 5.13, change \( y(x) \) to \( g(x) \).
Page 70, line 8, delete the sentence "A sampling estimate of a population parameter is called a statistic".

Page 70, line 12, delete (statistics) and insert after the question mark: A function of the observations used to estimate a population parameter is called an estimator. The numerical value obtained by using the estimator is called an estimate.

Page 70, line 14, change sample estimates to estimators.

Page 71, lines 8, 9, 11, 13, change statistics to estimators.

Page 71, line 7b, omit (statistics); line 4b, change statistics to estimators.

Page 73, lines 10b, 8b and 2b, change μ₁₁ to μ₁ ; line 5b, change σ₁² to σ₂² ; line 5b, the first bracket should be squared; lines 2b and 1b change μ₁ to μ.

Page 74, line 2, change σ₁² to σ₂² . Example 6.1, the right side of the formula for l should have \( \frac{1}{n} \) as a factor, also change μ₁ to μ, and a random sample to any.

Page 74, line 8, change sample estimate to estimator; lines 9 and 10, change estimate to estimator. Line 12, change μ₁ to μ.

Page 74, Example 6.2, change μ₁₁ to μ₁ and μ₁ to μ.

Page 75, Example 6.3, change all μ₁ to μ. Section 6.6, change all μ₁₁ to μ₁ and μ₁ to μ.

Page 77, line 9, change l₁₁ to l₁. Section 6.7, change μ₁₁ to μ₁.

Page 78, line 2, the 11 should be 1.

Exercises, Page 80-81. Re-number the exercises, i.e. change 6.1 to 6.2, etc.

New Exercise 6.4. Change \( X_i \) to \( X_{ii} \).

New Exercise 6.8. Change \( T_j \) to \( T_3 \). (a) change first sentence to: Suppose \( X_{ij} \) is \( N(μ_i, σ^2) \) where \( μ_i \) is the mean effect of the \( i \)th treatment. After second sentence add: Hint: \( E(T_i) = μ_i \). Four pertinent independent forms are: \( l_1 = T_0 + T_1 + T_2 + T_3 \), \( l_2 = -3T_0 + T_1 + T_2 + T_3 \), \( l_3 = T_3 - T_1 \), etc.
Page 82, lines 8 and 12 change estimates to estimators.

Page 83, line 3, change $< n <$ to $< u <$ . Line 8b change $\sqrt{n}$ to $n$.

Page 84, lines 4 and 5, change $x^2$ to $x_0^2$ . Line 6, change the period to a comma after Incomplete Gamma Function and add $1(u,p)$ . Line 7, change $x^2$ and $V$ to $u = \frac{x_0^2}{2(p+1)}$, and $v = \frac{v}{2} - 1$.

Page 84, line 1b, change $\int \left( \frac{d}{2} \right)$ to $\int \left( \frac{h}{2} \right)$.

Page 85, line 4, change $2V$ to $\sqrt{2V}$ . Line 11, change $i \geq 2$ to $i > 2$.

Exercise 7.5, change $r_1$ to $r$.

Page 88, line 14, place question mark after sentence. In heading for Section 7.7 delete estimates.

Page 93, line 7b, change $x^{11}$ and $x^{21}$ to $\{x^{11}\}$ and $\{x^{21}\}$ respectively.

Page 94, line 11b, change $m_1$ and $m_2$ to $u_1$ and $u_2$ respectively.

Page 95, line 9, change $x_2$ to $\overline{x}_2$.

Page 96, Exercise 7.14, change degrees to degree.

Page 99, line 1, write $\left( \frac{x_{22}^2}{2} \right)_{22}^{u^2 - 1}$ , i.e. in parentheses.

Page 101, line 5, change us to use.

Page 102, Exercise 7.27, change $u^2$ to $u^p$.
CHAPTER I

Introduction

1.1 - What is statistics? Dis-satisfied with attempts at giving a precise formal definition of their subject, mathematicians have on occasion defined mathematics as being those professional activities engaged in by mathematicians. Later in this chapter an attempt will be made to give a more formal definition of the subject of statistics. For the present, however, if in the above definition the words "mathematics" and "mathematician" are replaced by "statistics" and "statistician", a definition of statistics might be written as follows:

Statistics comprises those professional activities engaged in by statisticians.

In order to gain some insight into the nature of statistics then, it remains to present in some detail the steps taken by a statistician in some simple yet representative investigation.

1.2 - A representative investigation. The problem is: Do insects go in crowds, independently, or avoid each other?
Step (a). An hypothesis, sometimes referred to as a null hypothesis, is set up. In this case it is assumed that insects move independently of each other.
Step (b). An experiment is designed to test the hypothesis. Four insects are placed at the center of a circular disc which has been divided by two perpendicular diameters into four quarters. The insects crawl around in the disc and finally all crawl off the rim.
Step (c). Pertinent data are collected and tabulated. The departure from the disc is observed using the following notation:

- $\frac{1}{4}$, all leave in one quarter
- $3,1$, three leave in one quarter, one in another
- $2^2$, two leave in one quarter, two in another
- $2,1^2$, two leave in one quarter, one each in two others
- $\frac{1}{4}$, one each leaves from each quarter.

The first two columns of Table 1.1 give the observations made.

Table 1.1

Observations and Computations for Analysis

<table>
<thead>
<tr>
<th>1 Method of Departure</th>
<th>2 Observed Frequency</th>
<th>3 Expected Frequency</th>
<th>4 Contributions to Chi-square</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0</td>
<td>1.9</td>
<td>1.90</td>
</tr>
<tr>
<td>3,1</td>
<td>25</td>
<td>22.5</td>
<td>.28</td>
</tr>
<tr>
<td>$2^2$</td>
<td>11</td>
<td>16.9</td>
<td>2.06</td>
</tr>
<tr>
<td>$2,1^2$</td>
<td>73</td>
<td>67.5</td>
<td>0.45</td>
</tr>
<tr>
<td>$\frac{1}{4}$</td>
<td>11</td>
<td>11.2</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>120</strong></td>
<td><strong>120.0</strong></td>
<td><strong>4.69</strong></td>
</tr>
</tbody>
</table>

Step (d). The parent population distribution is specified. In this case this amounts to specifying that the probability that any insect leaves by any designated quarter is $1/4$. As will be noted later, specification of the probability of the occurrence of an observation will be done by stating that the observation or observations follow some parent population distribution such as the normal, binomial, or Poisson.
Step (e). The distribution of the data and for functions of the data, on the assumption that the hypothesis is true, are obtained. In this case the expected values in column 3 of Table 1.1 are obtained from corresponding values in Table 1.2 as explained below.

### Table 1.2

<table>
<thead>
<tr>
<th>1 Methods of Departure</th>
<th>2 Number of ways Group can be Formed</th>
<th>3 Number of ways Group can be Distributed</th>
<th>4 Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>3, 1</td>
<td>4</td>
<td>12</td>
<td>48</td>
</tr>
<tr>
<td>2, 2</td>
<td>3</td>
<td>12</td>
<td>36</td>
</tr>
<tr>
<td>2, 1, 1</td>
<td>6</td>
<td>24</td>
<td>144</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>1</strong></td>
<td><strong>256</strong></td>
<td></td>
</tr>
</tbody>
</table>

Column 2 of Table 1.2 is filled in the following manner. Let a, b, c, d represent the four insects, then we can form only one group with a, b, c, d. Corresponding to 3, 1 we have the groups: abc, d; abd, c; acd, b; and bcd, a; giving a total of 4. For 2, 2 we have ab, cd; ac, bd; ad, bc giving 3. For 2, 1, 1 we have ab, c, d; ad, b, c; bd, a, d; bd, a, c; cd, a, b making a total of 6. For 1, 1 we have a, b, c, d giving 1.

Column 3 of Table 1.2 is filled in as indicated below. The group of 4 may be distributed in the four quarters in 4 ways. The group 3, 1 may be distributed 4 x 3 = 12 ways; 2, 2 in 4 x 3 = 12 ways; 2, 1, 1 in 4 x 3 x 2 = 24 ways; 1, 1 in 4 x 3 x 2 x 1 = 24 ways.
Column 4 of Table 1.2 gives the numerators of the probabilities of getting 4, 3, 1, etc. In order to fill in column 3 of Table 1.1, we take \( \frac{4}{256} \times 120 = 1.9 \), etc., giving the expected values.

Step (f). A test of the hypothesis is made. In this case \( \chi^2 \) (4) defined as
\[
\frac{S(0-e)^2}{e} = 1.90 + .28 + \ldots 0.00 = 4.69 \text{ with 4 degrees of freedom}
\]
where \( O \) corresponds to an observed value and \( e \) its expected value, is calculated as the test criterion. From the available tables of the \( \chi^2 \) distribution (4) it is found that the probability of obtaining a \( \chi^2 \) this large or larger on the assumption that the null hypothesis is true is greater than .30. Since this is not an unreasonable value we say that the hypothesis is not refuted, i.e. from this experiment we have no evidence to contradict the hypothesis that insects move independently of each other.

The representative investigation described above is concerned with an experiment. The second large class of investigations involving the use of statistical methodologies is that of the sample survey. In general a sample survey provides estimates of population parameters and confidence limits while an experimental investigation usually leads to a statistical test of an hypothesis as well as estimates of population parameters and confidence limits.

Close inspection of the steps taken by the statistician in the above representative investigation will reveal the fact that most are similar or identical with those taken by workers engaged in many different fields of scientific enquiry. What essential characteristic then is peculiar to the professional activities of the statistician?
A careful comparison will reveal that this essential characteristic is the use of the mathematics of probability to calculate from the observations themselves a measure of the fallibility of conclusions and estimates. In the experiment this takes the form of a statistical test of hypothesis at an assigned probability level, while in a sample survey confidence limits are calculated similarly by the use of sample probability distribution tables. However, valid and efficient measures of the fallibility of conclusions in terms of exact probability statements are possible only if the earlier steps in the investigation are taken with this end product in mind. Hence, the statistician finds himself vitally concerned with matters not strictly concerned with this main aspect such as: statement of the null hypothesis, design of experiments and sample surveys, questionnaire construction and training and supervision of enumerators, experimental techniques, collection and tabulation of data, specification of the parent population distribution, and interpretation of results.

1.3 - A formal definition of statistics. With the above discussion in mind, statistics might be given the following formal definition:

Statistics is the science and art of the development and application of the most effective methods of collecting, tabulating, and interpreting quantitative data in such a manner that the fallibility of conclusions and estimates may be assessed by means of inductive reasoning based on the mathematics of probability.

1.4 - Probability. It was stated in the definition that statistics uses inductive reasoning based on the mathematics of probability.

In this respect then statistics is a branch of applied mathematics
whose methodologies stem from the axioms and theorems of probability which in turn is a branch of pure mathematics. A definition of the probability of the occurrence of an event would appear to be in order. Unfortunately there is no general agreement among workers in the field as to what constitutes a satisfactory definition. For reasons of simplicity the classical definition will be given which is as follows:

If an event can occur in \( N \) equally likely and mutually exclusive ways, and if \( n \) of these ways have an attribute \( A \), then the probability of the occurrence of \( A \) is \( n/N \).

This is the definition of a \textit{a priori} probability, that is, it assumes that it is possible to logically determine, before trials are made, all the equally likely and mutually exclusive ways that an event may happen and assign \( n \) of these ways to the occurrence of attribute \( A \).

\textbf{Example 1.1} - What is the probability of obtaining a head with a penny on a single toss. Assuming the coin a "true" coin we reason that it may fall two equally likely ways and that one of them must be heads, hence, the probability is \( 1/2 \).

Notice that the classical definition of probability, in using the words "equally likely", assumes a knowledge of probability in order to define the term. Logically, of course, this is certainly undesirable, but a more satisfactory definition must await a higher level of mathematical maturity than that assumed for this text.

In actual practice it would appear that \textit{a priori} probability might have limited usefulness. It may not be possible in a great many
important scientific problems to logically determine, before trials are made, all the equally likely and mutually exclusive ways that an event may happen and to assign a of these ways to the occurrence of attribute A. For instance, even in the example, the penny may have a tendency to turn up heads more often than tails, and the probability of heads, then, is no longer 1/2. But, what is the probability of heads now? Suppose the penny is tossed 100 times and 55 heads are noted. The probability of a head might be tentatively set as .55. However, we have only an estimate based on 100 tosses of some unknown postulated "true" probability in a theoretical infinite population of throws. This estimate is called the empirical probability.

What then is the connection between a priori probability and the sample estimate of some unknown hypothetical population probability, i.e. the empirical probability? With the a priori definition and certain postulates the mathematics of probability provides fundamental laws or theorems of probability which in turn make possible the solution of many classes of problems. If the unknown hypothetical population probability and its sample estimates of empirical probability be assumed to be amenable to the same fundamental laws, then a means becomes available for solving many important problems in the empirical sciences.

Exercise 1.1 - Calculate all expected values in column 3 of Table 1.1 and check the value of $X^2$ as 4.69.

Exercise 1.2 - After reading Reference 4 show that the probability of obtaining a $X^2$, with 4 degrees of freedom, as large or larger than 4.69 is greater than .30.
Exercise 1.3 - What is the \textit{a priori} probability of obtaining a seven with a pair of ordinary dice, if the dice are assumed to be "true"? Assuming that the pair of dice are not "true", how could one obtain a reasonable estimate of the unknown probability of obtaining a seven, i.e. the \textit{empirical} probability?

Exercise 1.4 - Read References 1, 2, and 3.

References and Other Reading


2. _______ \textit{The Design of Experiments}. Oliver and Boyd. Edinburgh. 1937. Chapters I and II.


CHAPTER II

Probability

2.1 - The not so "dry bones" of statistics. It was stated earlier that statistics concerns itself with inductive reasoning based on the mathematics of probability. In developing statistical methodology the statistician makes use of the definitions, postulates, and theorems of mathematical probability. What can be said of this framework, these "bones", of statistics? The theory of probability had its genesis in the application of mathematics to determining the odds in various games of chance: dice, cards, spun wheels, etc. In particular, the foundations of the science of probability was laid by two seventeen century mathematicians, Pascal and Fermat, in their private correspondence concerning questions raised regarding the gambling observations of the French nobleman, Chevalier de Méré. Books on games of chance are still being written by workers in probability and statistics (1).

Statistics is then no "dry as dust" subject concerning itself with the compilation of innumerable tables and charts. On the contrary, it deals with the development and application of an important methodology based on the fascinating subject of probability. This methodology has become of great importance as a research tool in the physical, biological, and social sciences.

2.2 - Number of ways an event can occur: permutations and combinations.

The number of equally likely and mutually exclusive ways that an event may occur may be determined by enumeration or by the use of some simple
rules from college algebra. The latter method is simpler for more complicated problems. Two fundamental theorems are:

**Rule 2.1.** If A can happen in \( m \) ways and B in \( n \) other ways independent of \( m \) then both A and B can happen in \( mn \) ways.

**Example 2.1.** If two ordinary dice are tossed, one may appear face up in 6 ways which are independent of the 6 ways that the second may appear. Hence, both may appear face up together in 36 different ways.

**Rule 2.2.** If A can happen in \( m \) ways and B in \( n \) other ways mutually exclusive of \( m \) then either A or B can happen in \( m + n \) ways.

**Example 2.2.** An ace or a king may be drawn from an ordinary deck of cards in \( 4 + 4 = 8 \) ways.

For multiple arrangements, a rule for permutations can be applied. If it is desired to arrange \( n \) different objects into sets of \( r \) objects per set, the number of such arrangements is called "the number of permutations of \( n \) objects taken \( r \) at a time" and is indicated by \( \frac{n!}{(n-r)!} \). The first of the \( r \) positions can be filled in \( n \) ways, the second in \( n-1 \) ways since one object will have been used in the first position, the third in \( n-2 \) ways, etc. Hence, by Rule 2.1,

**Rule 2.3.** \( n^r = n(n-1)(n-2)...(n-r+1) = \frac{n!}{(n-r)!} \).

where \( n! = n(n-1)...2.1 \).

It should be noted that, when \( r = n \), \( n^r = n! \) which implied that \( 0! = 1 \).

**Example 2.3.** The number of different ways of selecting a president, vice president, and secretary from a suggested slate of 6 is \( 6^3 = 6.5.4 = 120 \).
Suppose that all \( n \) objects in the arrangement are used, but certain groups \( n_1, n_2, \ldots \) are alike. Any re-arrangement of the objects of any \( n_1 \) group will not change any particular arrangement, hence, the total number of arrangements will be less than if all the objects were different from one another. Now, any group of \( n_1 \) alike objects can be arranged \( n_1! \) ways and, since these \( n_1! \) arrangements are alike for every arrangement of the other objects, the total number of different arrangements will be given as below.

\[
\text{Rule 2.4. } n^P(n_1, n_2, n_3, \ldots) = \frac{n!}{n_1! n_2! n_3! \ldots},
\]

where \( n^P(n_1, n_2, n_3, \ldots) \) represents the total number of permutations, given that \( n_1 \) are alike, \( n_2 \) alike but different from the first group, etc., and \( n = n_1 + n_2 + \ldots \).

**Example 2.4** - How many different 6-flag signals may be made if 3 are red, 2 blue, and one yellow? Answer: \( 6^P(3, 2, 1) = \frac{6!}{3!2!1!} = 60 \).

If interest lies only in groups of objects and not in the arrangements within the groups, then combinatorial rules apply. The total number of combinations of \( n \) objects taken \( r \) at a time is denoted symbolically as \( n^C_r \). It is easily seen that

\[
n^P_r = n^C_r \cdot r^P_r,
\]

since each combination of \( r \) objects may be permuted \( r \) times. The following rule is now derived.

**Rule 2.5** - \( n^C_r = \frac{n^P_r}{r^P_r} = \frac{n!}{(n-r)! r!} \)

**Example 2.5** - The total number of different bridge hands of 13 cards which can be dealt from a deck of 52 cards is \( 52^C_{13} = \frac{52!}{13! \cdot 39!} \).
Again the total number of sets of 4 bridge hands is
\[ 52 \binom{13}{4} \cdot 39 \binom{13}{2} \cdot 26 \binom{13}{2} \cdot 13 \binom{13}{1} = \frac{52!}{(13!)^4} \cdot \]

2.3 - Stirling's approximation. The use of the rules of permutations and combinations involves factorials, some with quite large values of \( n \). Stirling's formula,
\[ n! \sim \sqrt{2\pi n} \cdot n^{n-\frac{1}{2}} \cdot e^{-n} \}

may be used to obtain quickly an approximation to \( n! \). The first term,
\[ n! \sim \sqrt{2\pi n} \cdot n^{n-\frac{1}{2}} \]
gives a suitable approximation in many cases.

Example 2.6 - Evaluate \( 13! \) by the use of Stirling's formula,
\[ 13! \sim \sqrt{2\pi(13)} \cdot 13^{13-\frac{1}{2}} \cdot e^{-13} \cdot \left(1 + \frac{1}{12(13)^2}\right) \]
\[ \log 13! \sim \frac{1}{2} \left( \log 26 + \log n \right) + 13 \log 13 - 13 \log e + \log \frac{157}{156} \]
\[ 13! \sim 6.2271 \times 10^9 \]
using 5 place logarithm tables.

2.4 - Probability and arrangements. After obtaining the total number of mutually exclusive and equally likely ways and those that possess attribute \( A \), by the use of the rules of permutations and combinations, it is then possible to write the required probability by applying the fundamental definition
\[ p = \frac{\text{ways that possess attribute } A}{\text{total number of ways}} \]

Example 2.7 - A bag contains \( \frac{1}{2} \) red and 3 white balls. What is the probability of obtaining exactly 3 red balls? Answer:
\[ p = \frac{3^1}{7^3} = \frac{4}{35} \]

2.5 - **Fundamental laws of probability.**

**Law 1.** If A and B are two mutually exclusive events (the occurrence of one precludes the occurrence of the other) then the probability of either of them happening is the sum of the respective probabilities. Symbolically, \( P(A + B) = P(A) + P(B) \).

**Example 2.8** - The probability of throwing either a 7 or 8 with two dice is \( 6/36 + 5/36 = 11/36 \).

**Law 2.** If A and B are two independent events, so that the occurrence of one does not affect the chance of the occurrence of the other, the probability that both happen is the product of their respective probabilities. Symbolically, \( P(AB) = P(A) \cdot P(B) \).

**Example 2.9.** The probability of getting 2 red balls in drawing one ball from each of two urns containing 6 red balls and 1 black ball is \( 6/10 \cdot 6/10 = 9/25 \).

If A and B are not independent of one another so that the occurrence of one affects the probability of the others occurrence, then a definition is needed for the probability of A given that B has happened. This is the conditional probability of A and is denoted \( P(A|B) \). Similarly the conditional probability of B given that A has happened is \( P(B|A) \). In such cases fundamental law 2 becomes:

**Law 3.** \( P(AB) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B) \). If A and B are independent, \( P(B|A) = P(B) \) and \( P(A|B) = P(A) \).
Example 2.10 - If both balls were drawn in succession from one of the urns in Example 2.9 without replacement, then the probability of obtaining 2 red balls is $\frac{6}{10} \cdot \frac{5}{9} = \frac{1}{3}$.

Law 4. If two events are not mutually exclusive, then the probability of at least one of them occurring is $P(A+B) = P(A) + P(B) - P(AB)$.

Proof:
Let $\overline{A}$ and $\overline{B}$ represent the non-occurrence of $A$ and $B$ respectively. Then,

$P(A) + P(B) = P(AB) + P(\overline{A}B) + P(A\overline{B}) + P(\overline{A}\overline{B}) = P(AB) + P(A + B)$.

$\therefore P(A + B) = P(A) + P(B) - P(AB)$, where $P(\overline{AB})$ is the probability of $A$ occurring and $B$ not occurring, and similarly for $P(\overline{BA})$.

Example 2.11 - In Example 2.9 the probability of obtaining at least one red ball is

$$P(A + B) = P(A) + P(B) - P(AB)$$

$$= \frac{6}{10} + \frac{6}{10} - \frac{6}{10} \cdot \frac{6}{10} = \frac{21}{25}.$$ 

Law 5. If the probability of an event occurring in a single trial is $p$, the probability of its occurring $r$ times out of $n$ trials is given by

$$nC_r p^r (1-p)^{n-r} = nC_r p^r q^{n-r},$$

where $1 - p = q$.

Proof: If the event occurs $r$ times out of $n$ trials, it will fail to occur $n-r$ times; hence, the probability of the occurrence of any sequence of $r$ successes and $n-r$ failures is $p^r (1-p)^{n-r}$. But, the number of possible sequences is given by $nC_r$.

Example 2.12 - The probability of obtaining exactly 3 heads on a single toss of 5 coins is $5C_3 (1/2)^3 (1/2)^2 = 5/16$. 
Exercises

Exercise 2.1 - An agronomist is designing an experiment involving the use of 4 varieties, 3 fertilizers, and 3 spacings. How many different treatment combinations, using one from each of the three kinds of treatments, does he have?

Exercise 2.2 - In how many different ways may a Jersey or a Holstein be drawn from a mixed herd of 5 Jerseys, 7 Holsteins, 10 Guernseys, and 6 Brahamins?

Exercise 2.3 - How many different ways may a horticulturist arrange 5 different potted plants on a greenhouse bench?

Exercise 2.4 - How many different ways may a student select a major and a minor from 5 possible fields?

Exercise 2.5 - How many different arrangements can be made using the 10 letters from the word statistics?

Exercise 2.6 - How many signals can be made by hoisting 6 flags of different colors one above the other, when any number of them may be hoisted at once?

Exercise 2.7 - An organism has the possibility of having 1, 2, 3, 4 or 5 out of a total of 15 characters. What are the total possible combinations?

Exercise 2.8 - An industrial engineer in designing an experiment arranged to measure sources of variation from 4 factors (runs 4, journeys 5, cylinders 3, and pots 2). If we let $R_i (i = 1 \text{ to } 4)$, $J_j (j = 1 \text{ to } 5)$, $C_k (k = 1 \text{ to } 3)$, and $P_l (l = 1 \text{ to } 2)$ represent the respective factors, how many first order interactions of the type $R_i J_j$ etc. are there? How many second order interactions? How many third order interactions?
Exercise 2.9 - Using the relationship
\[(1+x)^n = 1 + \binom{n}{1}x + \binom{n}{2}x^2 + \ldots + \binom{n}{n-1}x^{n-1} + \binom{n}{n}x^n\]
show that
\[2^n - 1 = \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n-1} + \binom{n}{n}.

How many ways can we make a selection of 5 breeds of chickens?

Exercise 2.10 - Show that \(\binom{n}{r} = \binom{n}{n-r}\).

Exercise 2.11 - If \(\binom{n}{10} = \binom{n}{6}\), find \(\binom{n}{3}\).

Exercise 2.12 - If \(16^2 = 16^r - 2\), find \(r\).

Exercise 2.13 - If \(\frac{\binom{50}{r}}{\binom{51}{51-r}} = 30600\), find \(r\).

Exercise 2.14 - A random sample of size \(n\) from a finite population of \(m\) sampling units is one in which every possible combination of size \(n\) has an equal chance of being chosen. How many different samples of size 10 may be drawn from a list of 100 names? Use Sterling's approximation to evaluate the factorials.

Exercise 2.15 - Suppose that in selecting the sample of size 10 in Exercise 2.14 we draw a number from 1 to 10 at random, say 6, and select every 10th name on our list thereafter, i.e. 16, 26, etc. Is this method of selection equivalent to random sampling?

Exercise 2.16 - From a pack of 52 cards two are drawn at random; find the probability that one is a queen and the other a king. Ans. 8/663.
Exercise 2.17 - There are three events A, B, C, one of which must, and only one can, happen; the probabilities of A or B not happening are 5/11 and 5/7 respectively. Find the probability of C happening. Ans. 34/77.

Exercise 2.18 - The probability of A solving a certain problem is 3/7 and the probability of B solving the same problem is 5/12. What is the probability that the problem will be solved if both try? Ans. 2/3.

Exercise 2.19 - In a family with 6 children, what is the probability that:
(a) all children will be girls; (b) all children will be of the same sex;
(c) the first 5 children will be boys and the sixth a girl? (d) that 3 of the children will be boys? Assume the sex ratio is 1/2. Ans. (a) 1/64, (b) 1/32, (c) 1/64, (d) 20/64.

Exercise 2.20 - Show that, if the probability of an event happening in one trial is p and not happening is q, then the probability of the event happening exactly r times in n trials is the (r+1)th term in the expansion of $(q+p)^n$, that is $\binom{n}{r} p^r q^{n-r}$.

Exercise 2.21 - Show that under the conditions of Exercise 2.20 that the probability that an even happens at least r times in n trials is
$$p^n + \binom{n}{1} p^{n-1} q + \binom{n}{2} p^{n-2} q^2 + \ldots + \binom{n}{n-r} p^{r} q^{n-r},$$
or the sum of the first n-r+1 terms of the expansion of $(p+q)^n$.

Exercise 2.22 - A lady declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup. Eight cups of tea were mixed, four in one way and four in the other, and the lady so informed. The cups were then presented, in
random order, to the subject for judgment. The subject was asked to divide the 8 cups into two sets of 4, agreeing, if possible, with the treatments received. The lady selected 3 right and 1 wrong for one designated set of 4. On the assumption that the null hypothesis is true, i.e. the lady cannot discriminate between the two methods, show that the probability of the subject doing as good or better is $17/70$. Hence, we do not have evidence that the null hypothesis is false, if 1 in 20 be taken as a probability level for significance.

References and Other Reading


Chapter III

Univariate Parent Population Distributions

3.1-Specification. In order to arrive at the end-product, a test of an hypothesis in the representative statistical investigation discussed in Chapter I, it was specified that the probability that any insect leaves by any particular quarter is \( \frac{1}{4} \) and is independent of the probability for any other insect. Essentially this means that the probabilities of a complete set of chance events are given, instead of the probability of a single chance event as discussed in Chapter II. In this case, we say that the parent population probability distribution is given by specifying the following table of values:

<table>
<thead>
<tr>
<th>( X )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f(X) )</td>
<td>( \frac{4}{256} )</td>
<td>( \frac{48}{256} )</td>
<td>( \frac{36}{256} )</td>
<td>( \frac{144}{256} )</td>
<td>( \frac{24}{256} )</td>
</tr>
</tbody>
</table>

Notice that the descriptive occurrences \( \frac{4}{256}, \frac{3}{256}, \frac{2}{256}, \frac{1}{256}, \) and \( \frac{4}{256} \) have been coded to 1, 2, 3, 4, 5 respectively. Ordinarily, in applied statistics, specification is accomplished by choosing a mathematical form or function, e.g. the normal binomial, or Poisson, and simply stating that the observations are a random sample of all possible values of \( X \). Quoting from R. A. Fisher (1):

"... we may know by experience what forms are likely to be suitable, and the adequacy of our choice may be tested a posteriori. We must confine ourselves to those forms which we know how to handle, or for which any tables which may be necessary have been constructed."
3.2 - **Discrete distributions.** Functions like \( f(x) \) in Table 3.1 are called discrete probability distribution functions to distinguish distributions of this type from continuous probability distribution functions to be discussed later. The various values of \( f(x) \) may be thought of as giving the relative frequencies of occurrence corresponding to the particular values of \( x \). In the illustrative investigation 120 trials were made with four insects, hence, the expected values or theoretical frequencies were obtained for Table 1.1 by taking \( \frac{4}{256} \times 120 = 1.9 \), etc.

The discrete distribution given by Table 3.1 may be exhibited graphically as in Figure 3.1 below:

![Figure 3.1](image)

*Figure 3.1. Graph of Discrete Probability Distribution of Table 3.1 (f(x) axis given in parts of 256)*

Since some one of the five events must occur on any one trial, the sum of all the probabilities is one, or symbolically

\[
\sum_{x=1}^{f} f(x) = 1.
\]  

The distinguishing characteristic of the discrete distribution is that the variate \( x \) can take only isolated values, e.g. in Figure 3.1 only the whole numbers 1 through 5.
3.3 - The binomial distribution. This discrete distribution is the distribution of successes \( x \) in repeated independent trials \( n \) in which the probability of success on any trial is a constant \( p \). It has been named the binomial distribution because the successive probabilities are given by the respective terms of the expansion of the binomial \((p+q)^n\), where \( q = 1 - p \). One property of the binomial theorem is that the \((x+1)\)st term of the expansion is

\[
f(x) = \binom{n}{x} p^x q^{n-x}
\]

which by the methods of Chapter II also gives the probability of exactly \( x \) successes in \( n \) trials. On the right side of (2) \( x \) is the variate and \( p \) and \( n \) are parameters, i.e. for any particular number of this family of distributions special values of \( p \) and \( n \) must be specified. Since

\[
\sum_{x=0}^{n} f(x) = (p+q)^n
\]

this distribution fulfills the requirement that the sum of the probabilities is one.

Example 3.1 - Given that the probability of drawing a tenant farm in a sample of North Carolina farms is \( 1/3 \). If samples of 5 farms are drawn, then the respective probabilities of obtaining 0, 1, 2, 3, 4, 5 tenant farms in a single sample are:

\[
(2/3)^5; 5(1/3)(2/3)^4; 10(1/3)^2(2/3)^3; 10(1/3)^3(2/3)^2; 5(1/3)^4(2/3); (1/3)^5
\]

or 1/243 \([32, 80, 80, 40, 10, 1]\).

The probabilities \( f(x) \) of obtaining \((x=0, 1, 2, 3, 4, 5)\) tenant farms may be shown graphically as in Figure 3.2.
Figure 3.2. Graph of Probabilities of Obtaining tenant farms 
(f(x) axis given in parts of 243)

If the probabilities are accumulated and graphed as in Figure 3.3, then 
some P(a) value gives the probability of obtaining a value of x less 
than or equal to a, i.e. P(a) = P(x ≤ a). This step function is called 
a cumulative distribution or an ogive.

Figure 3.3 Cumulative Distribution of Tenant Farm Probabilities 
(F(x) axis given in parts of 243)
Note that points of discontinuity occur for each whole number on the x axis and that \( P(5) = P(x \leq 5) = 1 \).

3.4 - The Poisson distribution. Another discrete distribution of importance in applied statistics is the Poisson distribution. The Poisson distribution may be derived as a limiting form of the binomial distribution when \( p \) is very small, but \( n \) is so large that \( np \) is a finite constant, equal to \( \mu \), say.

To see this consider the binomial distribution:

\[
g(x) = \frac{n(n-1)(n-2) \ldots (n-x+1)}{x!} p^x q^{n-x}.
\]

Since \( p = \frac{\mu}{n} \),

\[
g(x) = \frac{n(n-1)(n-2) \ldots (n-x+1)}{x!} \left( \frac{\mu}{n} \right)^x \left( 1 - \frac{\mu}{n} \right)^{n-x}
\]

\[
= \frac{(1 - 1/n)(1 - 2/n) \ldots (1 - (x-1)/n) \mu^x (1 - \mu/n)^{n-x}}{x!}.
\]

Then

\[
limit_{n \to \infty} g(x) = \frac{x^m e^{-m}}{x!},
\]

but, this function is also a function of \( x \), hence the Poisson distribution may be written as

\[
f(x) = e^{-m} \frac{\mu^x}{x!}, \quad 0 \leq x \leq \infty
\]

Since

\[
\sum_{x=1}^{\infty} \frac{\mu^x}{x!} = e^\mu,
\]

then

\[
\sum_{x=0}^{\infty} f(x) = 1.
\]

The distribution was named for Poisson, having been given first by him in 1837.

It should be noted that the Poisson distribution is a one parameter family, \( \mu \) being the parameter.
Example 3.2 - A bag of clover seed is known to contain 1% weed seeds.

A sample of 100 seeds is drawn. Since, \( m = np = 100(0.01) = 1 \), and \( e^{-1} = 0.3679 \), the probabilities of 0, 1, 2, 3, ..., weed seeds being in the sample are:

<table>
<thead>
<tr>
<th>Number of Weed Seed</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>.3679</td>
<td>.3679</td>
<td>.1838</td>
<td>.0613</td>
<td>.0153</td>
<td>.0031</td>
<td>.0005</td>
<td>.0001</td>
</tr>
</tbody>
</table>

3.5 - Continuous distributions. If measurements instead of counts constitute the data under consideration, then the hypothetical parent population distribution is that of a continuous variate instead of a discrete variate.

Snedecor (2) gives the following histogram for the gains in weight of 100 swine.

![Histogram showing frequency of gains in weight of 100 swine](image)

**Figure 3.4.** Histogram Showing Frequency of Gains in Weight of 100 Swine

Before powerful mathematical methods may be applied, in order to derive a methodology providing techniques for statistical inferences, it is desirable to "idealize" the histogram into a curve which may be represented by a mathematical function. Such a process takes place in other branches of applied mathematics, e.g. in surveying. Before the surveyor can be furnished with
a powerful methodology for the solution of his practical problems in mensuration, it is necessary for the geometer to "idealize" the physical points, lines, and planes. A geometrical point is defined as having no dimensions, but simply an indicator of position. Again, a geometrical line has no width and a geometrical plane has no thickness. With these "idealized" definitions and certain assumptions called axioms the geometer is able to prove theorems concerning relationship and properties of geometrical configurations. These theorems in turn form the bases for a practical surveying methodology.

How shall we "idealize" a histogram of the type given in Figure 3.4? First of all, instead of a finite sample, we assume an infinite population of gains in weight. Next, instead of the class marks differing by 5 pounds, suppose that they are selected closer and closer together. It is not difficult to see that the histogram might reasonably be expected to approach some continuous smooth curve of the type shown in Figure 3.5.

In Figure 3.2 the probability of obtaining some particular $x$ for the discrete distribution was represented by an ordinate. Here, $x$ represents only whole numbers, but, because of the limitations imposed by measuring devices, the best that can be said concerning the "true" value of some observed $x$-value where $x$ represents a continuous variate is that it lies in some interval $(x, x+dx)$. If the area under the continuous curve be made equal to one, corresponding to the

![Figure 3.5. "Idealized" or Theoretical Probability Distribution](image-url)
similar requirement that the sum of the probabilities equal one for the discrete distributions, then the probability of \( x \) lying in the interval \((x, x+dx)\) will be \( f(x)dx \). This would be the theoretical probability corresponding to the empirical probability, say, of a gain of weight lying between \( 22.5 \) and \( 27.5 \), i.e. \( 13/100 = .13 \).

The range of \( x \) may be thought of as extending from \(-\infty\) to \(+\infty\), even though the curve may actually contact the x-axis at some finite value, since the area under the curve in the contact interval would be zero. As was pointed out the total probability or area under the curve is one, symbolically

\[
\int_{-\infty}^{\infty} f(x)dx = 1. \quad (3)
\]

The probability of \( x \) being equal to or less than some constant \( a \) is expressed as

\[
P(x \leq a) = \int_{-\infty}^{a} f(x)dx. \quad (4)
\]

Again, the probability of \( x \) lying between \( a \) and \( b \) is given by

\[
P(a \leq x \leq b) = \int_{a}^{b} f(x)dx. \quad (5)
\]

It is possible to omit the equal signs in the left sides of (4) and (5) since the probability of obtaining any particular value of \( x \) is equal to the width of a geometrical line which is zero.

The cumulative probability distribution or ogive for the continuous variate corresponding to the probability distribution of Figure 3.5 would appear as in Figure 3.6.

Figure 3.6. Cumulative Probability Distribution For a Continuous Variate
For this curve, $P(a) = P(x \leq a)$ and hence $P(b) - P(a) = P(a \leq x \leq b)$.

3.6. - The normal distribution. The most important continuous distribution in applied statistics is, of course, the normal distribution. The histogram of Figure 3.4 and the theoretical distribution of Figure 3.5 are those for data specified as being normally distributed. Data arising from many different measurements taken on plants and animals are specified as following the normal distribution. There is, of course, empirical justification for this assumption. Similarly distributions of certain data of the physical and social sciences are found to be satisfactorily represented by the normal distribution. It should not be assumed, however, that every continuous distribution representing actual data should be normal. For example, it is known that the distribution of sizes of cumulus clouds should be represented by a u-shaped curve. The mathematical form of the normal curve is defined by

$$f(x)dx = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-u)^2} dx, -\infty \leq x \leq \infty$$

The curve is symmetrical about $x = u$, and bell shaped as Figure 3.5. The inflection points are at $x = \pm \sigma$ and the tails of the curve, although approaching the x-axis quite rapidly, extend indefinitely far in both directions. The function (6) represents a two parameter family, $u$ and $\sigma$, of continuous distributions.

Since it can be shown for (6) that

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

then the normal probability distribution has this same property in common with the binomial and the Poisson distributions.
3.7 - **Probability distributions as specialized mathematical functions.**

We have noticed that theoretical probability distributions are mathematical functions possessing certain requirements. In order to give a complete formal definition of the requirements necessary for a mathematical function to be a probability distribution of statistics it is convenient and sufficient to consider the *cumulative distribution function*, $F(x)$. It is sufficient since, given the cumulative distribution, it is possible to find the distribution itself by taking the differential, i.e.

$$
\frac{d}{dx} \left[ F(x) \right] = f(x)dx.
$$

A mathematical function, $F(x)$, may be used as a cumulative distribution of a chance variable provided that

(a) $F(-\infty) = 0$, $F(\infty) = 1$

(b) $F(x)$ is a monotonically increasing function, i.e. if $x_1 > x_2$, $F(x_1) \geq F(x_2)$.

(c) $F(x)$ is defined at every point in a continuous range and is continuous, except possibly at a denumerable number of points.

The following notation should be kept clearly in mind:

- $f(x)dx$ is the probability distribution or distribution,
- $f(x)$ is the frequency function,
- $F(x)$ is the cumulative distribution.

3.8 - **Some mathematical functions useful as probability distributions.** Karl Pearson has suggested the differential equation

$$
\frac{dy}{dt} = \frac{(m-t)y}{at+b+ct^2}
$$

as a generator of possible parent population distributions useful in applied statistics. For example, if $m = b = c = 0$ and $a = 1$, then the differential equation becomes

$$
\frac{dy}{y} = -t \, dt.
$$
Solving for $y$ we have

$$\log_e y = -\frac{t^2}{2} + c_1.$$

Then

$$y = ce^{-\frac{t^2}{2}}.$$

Upon setting the integral between the limits from $-\infty$ to $\infty$ equal to 1, we find $c = \frac{1}{\sqrt{2\pi}}$. Letting

$$t = \frac{x - \mu}{\sigma},$$

we obtain the function

$$y = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty,$$

which is the normal frequency function. This is Type VII of the Pearson system of frequency functions (6).

Another method of obtaining a mathematical representation of a frequency function is furnished by the Gram-Charlier series (3).

3.9 - The Gamma and Beta Functions. These are two useful functions in statistics. The Gamma function of the positive number $n$ is defined by

$$\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}dx, \quad n > 0.$$
Exercise 3.1 - Using integration by parts to show that \( \Gamma(n+1) = n \Gamma(n) \).

Exercise 3.2 - Show that \( \Gamma(n+1) = n(n-1) \cdots (n-k) \Gamma(n-k) \) where \( k \) is a positive integer less than \( n \).

Exercise 3.3 - If \( n \) is also a positive integer, show that \( \Gamma(n+1) = n! \).

Exercise 3.4 - (a) Find: \( \Gamma(1), \Gamma(2), \Gamma(3), \) and \( \Gamma(4) \).

(b) Using \( \Gamma(n+c) = (n+c-1) \cdots (n+1)n \Gamma(n) \) show that \( \Gamma(n) \) becomes infinite when \( n = 0 \).

Exercise 3.5 - Show that

\[
\Gamma(n) = 2 \int_0^\infty y^{2n-1} e^{-y^2} dy
\]

by setting \( x = y^2 \) in the integral defining the Gamma function.

Exercise 3.6 - Using the result of Exercise 3.5 show that:

(a) \( \Gamma\left(\frac{1}{2}\right) = 2 \int_0^\infty e^{-y^2} dy \)

(b) \( \left[\Gamma\left(\frac{1}{2}\right)\right]^2 = 4 \int_0^\infty \int_0^\infty e^{-(x^2+y^2)} dx dy \)

(c) \( \left[\Gamma\left(\frac{1}{2}\right)\right]^2 = 4 \int_0^{\pi/2} \int_0^\infty e^{-r^2} r dr dg \) in polar coordinates.

(d) \( \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \).

The Incomplete Gamma function defined by

\[
\Gamma(x) = \int_1^x \frac{1}{t^{n-1}} e^{-t} dt, \quad 0 \leq x \leq \infty,
\]

\[
= \frac{\Gamma(x)}{\Gamma(n)}.
\]
furnishes a useful cumulative distribution function.

The Beta function, defined by

\[ B(m,n) = \int_0^1 x^{m-1}(1-x)^{n-1} \, dx \]

is also of importance in theoretical and applied statistics.

Exercise 3.7 - By setting \( x = \sin^2 \theta \) in the integral defining \( B(m,n) \) show that

\[ B(m,n) = 2 \int_0^{\pi/2} \sin^{2m-1} \theta \cos^{2n-1} \theta \, d \theta. \]

Exercise 3.6 - By setting \( x = 1 - y \) in the integral defining \( B(m,n) \) show that

\[ B(m,n) = B(n,m). \]

Exercise 3.9 - Show that

(a) \( \Gamma(n) \Gamma(m) = 4 \int_0^\infty \int_0^\infty x^{2n-1} y^{2m-1} e^{-(x^2+y^2)} \, dx \, dy \)

(b) \( \Gamma(n) \Gamma(m) = 4 \int_0^{\pi/2} \sin^{2m-1} \theta \cos^{2n-1} \theta \, d \theta \int_0^\infty x^{2(m+n)-1} \, e^{-x^2} \, dx \)

(c) \( \Gamma(n) \Gamma(m) = B(m,n) \Gamma(m+n) \) or

\[ B(m,n) = \frac{\Gamma(m) \Gamma(n)}{\Gamma(m+n)} \]

The cumulative Incomplete Beta function is defined by

\[ F(x) = I_x(m,n) = \frac{1}{B(m,n)} \int_0^x \left( 1-x \right)^{n-1} x^{m-1} \, dx, \quad 0 \leq x \leq 1. \]

Both \( I_x(n), (7) \), and \( I_x(m,n), (8) \), have been tabulated by Karl Pearson and his staff at the Biometric Laboratory, University College London.
Exercises

Exercise 3.10 - Construct a parent population probability distribution for the sum of numbers appearing when two dice are tossed. Give: (a) a table of \( x \) and \( f(x) \), (b) a graph of (a), and (c) the cumulative distribution graph.

Exercise 3.11 - Follow the instructions in Exercise 3.10 for the probability of runs of heads of various length in tossing six pennies.

Exercise 3.12 - If you have a set of random numbers from 0 to 999, how would you set up a sampling scheme to select a random sample of 50 from 490 farms, the farms being numbered 0 to 489?

Exercise 3.13 - Follow the instructions in Exercise 3.10 for the binomial distribution with \( p = \frac{2}{5}, n=5 \).

Exercise 3.14 - Follow the instructions in Exercise 3.10 for the Poisson distribution with \( p=0.002, n=1,000 \).

Exercise 3.15 - Given the frequency function \( f(x) = 2x, \ 0 \leq x \leq 1 \), \( f(x) = 0 \) for \( x < 0 \) or \( x > 1 \). Show that \( \int_0^1 f(x) dx = 1 \). What is the functional form of the cumulative distribution function?

Exercise 3.16 - Show that the area under the curve is 1 for the triangular distribution function

\[ f(x) = \frac{2}{b}(1 - x/b), \quad 0 \leq x \leq b. \]

Exercise 3.17 - Repeat Exercise 3.16 for the rectangular distribution

\[ f(x) = 1/w, \quad 0 \leq x \leq w. \]

Exercise 3.18 - Repeat Exercise 3.16 for the Cauchy distribution

\[ f(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad -\infty \leq x \leq \infty. \]
Exercise 3.19 - Repeat Exercise 3.16 for the negative binomial distribution

\[ p_x = q^{-k} \frac{(k + x - 1)!}{x!(k - 1)!} (p/q)^x, \]

where \( q - p = 1, \quad 0 \leq x \leq \infty \)

Exercise 3.20 - A random variable \( X \), which lies between the limits 0 and 10, has the density function \( f(x) = Ax^3 \). Determine the value of \( A \) so that the total probability is 1. What is the probability that \( X \) lies between 2 and 5, that \( X \) is less than 3?

Exercise 3.21 - A random variable follows the normal distribution. Determine the coordinates of the maximum point on the frequency curve and show that \( + \sigma \) are the inflexion points. Show that the area under the normal curve is 1.

Exercise 3.22 - A variate has the distribution \( f(x) = e^{-x} \) in the interval \( 0 \leq x \leq \infty \). The probability is 1/2 that \( x \) will exceed what value?

Exercise 3.23 - If \( P(x < x_1) = 1 - \frac{1}{1 + x_1^2} \), \( x \) being a continuous variate with range \( 0 \leq x \leq \infty \), find the frequency function \( f(x) \).

Exercise 3.24 - May \( f(x) = -\frac{1}{(x-2)^2} \), \( 0 \leq x \leq 4 \), serve as a probability distribution function? Why?

Exercise 3.25 - Use the normal integral table to determine, for the normal distribution given in Exercise 3.21, the probability of: (a) \( x \geq \mu \).

(b) \( (\mu - \sigma) \leq x \leq (\mu + \sigma) \), (c) \( x \geq \mu + 2\sigma \).

Exercise 3.26 - Give the Poisson approximation to the binomial distribution with \( n = 2048 \) and \( p = 1/1024 \). Hence, obtain the probabilities of there being 0, 1, 2, 3, ..., times that ten tails appear in 2048 tosses of ten coins.
Exercise 3.27 - It can be shown for large $n$ that the binomial distribution may be approximated by the normal distribution with $\mu = np$ and $\sigma = npq$. If 20 coins are tossed, obtain the approximate probability of obtaining 19 or more heads.

Exercise 3.28 - For the distribution $f(x) = 2x$, $0 \leq x \leq 1$, find the number $a$ such that the probability of $x \geq a$ is 3 times the probability of $x \leq a$.

Exercise 3.29 - If two values of $x$ are drawn at random from the distribution $f(x) = e^{-x}$, $0 \leq x \leq 1$, what is the probability that both are greater than 1?

Exercise 3.30 - In Exercise 3.29 what is the probability that at least one value of $x$ is greater than 1?

Exercise 3.31 - In a certain species a plant with a fern type leaf is crossed with a plant having a palm type leaf and a family of 12 plants grown from the seed so obtained. Of the 12 progeny 2 had fern leaves and 10 palm leaves. Now fern is always homozygous $ff$, while palm may be $Ff$ or $FF$. Does the result of the experiment agree with the hypothesis that the cross was $(Ff + ff)$, i.e., the expectation of the progeny is 1:1? Hint: use the binomial distribution $(q+p)^n$. 
References and Other Reading


Chapter IV

Properties of Univariate Distribution Functions

4.1 - Introduction. In this chapter certain important properties of parent population distributions will be discussed. The discussion will also apply to similar properties of derived sampling distributions. These properties will be found useful in describing parent population distributions and derived sampling distributions.

4.2 - Mathematical expectation. The mathematical expectation of any random variable \( x \), which can assume values \( x_1, x_2, \ldots, x_n \) with probabilities \( p_1, p_2, \ldots, p_n \), respectively, where \( \sum_{i=1}^{n} p_i = 1 \), is defined to be:

\[
E(x) = \sum_{i=1}^{n} x_i p_i
\]

For the discrete distribution this becomes

\[
E(x) = \sum_{i=1}^{n} x_i p_i \]

since \( p_i = f(x_i) \), and for the continuous distribution, \( p_i = f(x_i)dx_i \), and with \( n \to \infty \),

\[
E(x) = \int_{-\infty}^{\infty} x f(x)dx
\]

where as noted before \( f(x) \) may vanish over part of the range \((-\infty, +\infty)\).

The term "mathematical expectation" may be shortened to "expected value" or "average value".

The definition of the expected value of any random variable \( x \) may be generalized to include functions of \( x \). The expected value of any function of \( x \), say \( \theta(x) \) is
\[ E[\Theta(x)] = \Sigma \Theta(x) f(x) \text{ or } \int \Theta(x) f(x) \, dx, \]

over the range of \( x \) and depending upon whether \( x \) is a discrete or continuous variate.

It is possible, by introducing a generalized form of summation called a "Stielje's integral", to replace the \( \Sigma \) or \( \int \) by a single integral sign. Although the concept of the Stielje's integral simplifies statements concerning probabilities and expected values, the mathematical concepts behind this refinement is beyond the scope of this text. A distinction in all statements will be made between the discrete and continuous distributions.

If the indicated integration, necessary to obtain the expected value, cannot be performed directly, various methods of numerical integration are available.

Example 4.1 - The expected value of \( x \), where \( x \) gives the various sums possible to be made by throwing two ordinary dice, may be found from the following table:

\[
\begin{array}{c|ccccccccccc}
  x & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \\
  36 f(x) & 1 & 2 & 3 & 4 & 5 & 6 & 5 & 4 & 3 & 2 & 1 & 1 \\
\end{array}
\]

Then,

\[
E(x) = \sum_{i=1}^{11} x_i \, f(x_i) = \frac{(2+6+12+\ldots+12)}{36} = 7.
\]

Example 4.2 - If \( f(x) = 3x^2 \), \( 0 \leq x \leq 1 \), then,

\[
E(x) = \int_0^1 x \cdot 3x^2 \, dx
\]

\[ = 3/14. \]
Also,

\[ E(x^n) = \int_0^1 x^n \cdot 3x^2 \, dx \]

\[ = \frac{3}{5}. \]

4.3 - Operations with expected values. The rules stated below will be found useful in operating with expected values.

1. The expected value of a constant is the constant itself. \( E(c) = c \).

2. The expected value of a constant times a variable is the constant times the expected value of the variable

\[ E[c \cdot \theta(x)] = cE[\theta(x)] \]

3. The expected value of a sum (or difference) is the sum (or difference) of the expected values of the separate parts.

\[ E[\theta_1(x) \pm \theta_2(x)] = E[\theta_1(x)] \pm E[\theta_2(x)] \]

The proofs of these statements are left as exercises for the student.

4.4 - Moments. The expected value of \( x^k \) is called the \( k^{th} \) moment of \( x \) about the origin and is represented by the symbol \( \mu_k \). Hence,

\[ \mu_k = E(x^k) = \Sigma x^k f(x) \text{ or } \int x^k f(x) \, dx, \]

over the range of \( x \). The first moment of \( x \) about the origin is referred to as the mean of \( x \), and is denoted by \( \mu_1 \). To simplify writing let \( \mu_1 = \mu \).

The expected value of \( (x-\mu)^k \) is called the \( k^{th} \) moment of \( x \) about the mean and is designated by the symbol \( \mu_k \). Hence

\[ \mu_k = E(x-\mu)^k = \Sigma (x-\mu)^k f(x) \text{ or } \int (x-\mu)^k f(x) \, dx, \]

over the range of \( x \). It is easily seen that

\[ \mu_1 = E(x-\mu) = 0. \]
The second moment about the mean, $\mu_2$, is called the \textit{variance} and is usually designated by the symbol $\sigma^2$. Hence,

$$\sigma^2 = \mu_2 = \mathbb{E}(x-\mu)^2 = \mathbb{E}(x^2) - \mu^2 = \int (x-\mu)^2 f(x) \, dx,$$

over the range of $x$. The formula for $\sigma^2$ may be written as follows:

$$\sigma^2 = \mathbb{E}(x-\mu)^2 = \mathbb{E}(x^2) - 2\mu \mathbb{E}(x) + (\mu)^2 = \mathbb{E}x^2 - (\mu)^2 = \mu_2 - (\mu)^2.$$The square root of $\sigma^2$ or $\sigma$ is referred to as the \textit{standard deviation}.

The third moment about the mean, $\mu_3$, furnishes a measure of \textit{skewness} or departure from symmetry about the mean of the distribution. One of the most generally accepted measures of skewness is

$$\alpha_3 = \frac{\mu_3}{\sigma^3}.$$A measure of the relative flatness or peakedness of the distribution, called the \textit{kurtosis}, is given by

$$\frac{\mu_4}{\sigma^4} - 3.$$Example 4.3 - Using the table of values of Example 4.1, we see that

$\mu = 7$. Also

$$\mu_2 = \frac{4+18+18+ \cdots + 144}{36} = \frac{1974}{36}.$$Hence,

$$\sigma^2 = \frac{1974}{36} - 49 = \frac{35}{6}.$$Example 4.4 - Using the distribution function of Example 4.2, i.e.,

$f(x) = 3x^2$, $0 \leq x \leq 1$:

$\mu = 3/4$, $\sigma^2 = 3/5 - 9/16 = 3/80$.

and

$$\mu_3 = \int_0^1 (x - 3/4)^3 \cdot 3x \, dx = -1/160.$$
Example 4.5. In order that the function,
\[ f(x)dx = ke^{-c(x-m)^2} \quad dx, \quad -\infty \leq x \leq \infty \]
represent a probability distribution, it is necessary that
\[ \int_{-\infty}^{\infty} f(x)dx = 1. \]
Hence, we find \( k = \frac{e}{\sqrt{2\pi}} \). Also,
\[ \mu = \int_{-\infty}^{\infty} x f(x)dx = m, \text{ and } \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = 1/2c. \]
Substituting these values in the original function we find
\[ f(x)dx = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad dx, \quad -\infty \leq x \leq \infty, \]
which is the normal distribution.

Example 4.6 - For the binomial distribution,
\[ f(x) = \binom{n}{x} p^x q^{n-x}. \]
\[ \mu = \sum_{x=0}^{n} x \frac{n!}{(n-x)!x!} p^x q^{n-x} \]
\[ = np \sum_{x=1}^{n} \frac{(n-1)!}{(n-x-1)!(x-1)!} p^{x-1} q^{n-x} \]
\[ = np(p+q)^{n-1} \]
\[ \therefore \mu = np. \]

4.5 - Moment-generating functions. The expected value of \( e^{tx} \) often provides a convenient short-cut in evaluating the moments of \( x \).
Since
\[ e^{tx} = 1 + \frac{tx}{1!} + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \ldots \]

\[ = \sum_{i=0}^{\infty} \frac{t^i}{i!} x^i \]

then
\[ E(e^{tx}) = \sum_{i=0}^{\infty} \mu_i \frac{t^i}{i!} \]

We set, say,
\[ \phi'(t) = E(e^{tx}) \]

and designate \( \phi'(t) \) the moment-generating function (m.g.f.) of \( x \). If \( \phi'(t) \) be differentiated \( k \) times with respect to \( t \) and then evaluated at \( t=0 \), we note that
\[ \left. \frac{\partial^k \phi'(t)}{\partial t^k} \right|_{t=0} = \mu_k. \]

To obtain the m.g.f. of \( x - \mu \) we consider
\[ e^{t(x-\mu)} = 1 + \frac{tx(x-\mu)}{1!} + \frac{t^2(x-\mu)^2}{2!} + \frac{t^3(x-\mu)^3}{3!} + \ldots \]

Then,
\[ E\left[ e^{t(x-\mu)} \right] = 1 + \sum_{i=1}^{\infty} \mu_i \frac{t^i}{i!} \]

If we set \( \phi(t) = E\left[ e^{t(x-\mu)} \right] \), then
\[ \mu_k = \left. \frac{\partial^k \phi(t)}{\partial t^k} \right|_{t=0} \]

In general the m.g.f. of any function, \( \theta(x) \), may be defined as
\[ E\left[ e^{t \theta(x)} \right]. \]
Example 4.7 - For the distribution,

\[ f(x)dx = e^{-x}dx \quad 0 \leq x \leq \infty. \]

\[ \phi(t) = \int_0^\infty e^{-(1-t)x}dx = 1 - t = \sum_{i=0}^\infty t^i. \]

Hence,

\[ u^i_k = k! \]

Example 4.8 - For the binomial distribution

\[ \phi(t) = \sum_{x=0}^n e^{tx} \binom{n}{x} p^x q^{n-x} \]

\[ = \sum_{x=0}^n \binom{n}{x} (p e^t)^x q^{n-x} \]

\[ = (p e^t + q)^n. \]

Then,

\[ \mu_1^i = np e^t(p e^t + q)^{n-1} \bigg|_{t=0} = np, \]

\[ = np \left[ e^{tn-1}(p e^t + q)^{n-2} p e^t + (p e^t + q)^{n-1} e^t \right]_{t=0} \]

\[ = np \left[ (n-1)p + 1 \right]. \]

Hence,

\[ \mu_2 = \sigma^2 = \mu_1^i - (\mu)^2 \]

\[ = np(1-p) \]

\[ = npq. \]

We may obtain the following relation between \( \phi(t) \) and \( \phi'(t) \):

\[ \phi(t) = E(e^{tx}) \]

\[ = e^{-t\mu} E(e^{t\gamma}) \]

\[ = e^{-t\mu} \phi'(t). \]
Beside furnishing a simple method of obtaining the moments in certain cases, the m.g.f. or more generally the characteristic function is of use in deriving distribution functions and in comparing distributions. These latter two uses follow from Theorems 4.1 and 4.2 of Section 4.5.

The expected value of $e^{tx}$ may not exist for real values of $t$ for many discrete and continuous distributions, e.g., for the Cauchy distribution

$$f(x)dx = \frac{1}{\pi} \frac{dx}{1 + x^2} \quad , \quad -\infty \leq x \leq \infty.$$

A more general function, which can be proved always to exist, is the characteristic function defined as $Ef(e^{itx})$ where $t$ is real. The characteristic function for the Cauchy distribution is $e^{-|t|}$. However, the evaluation of the integral necessary to obtain the characteristic function makes use of more advanced mathematical methods than are assumed for this course. Our uses of Theorems 4.1 and 4.2 will be confined to the m.g.f. which will be assumed to exist in such cases.

4.6 - Cumulants. Suppose that we define

$$\log \phi(t) = K(t) = K_1 t + K_2 \frac{t^2}{2!} + \ldots + K_r \frac{t^r}{r!} + \ldots$$

But,

$$\log \phi(t) = t\mu + \log \phi(t)$$

and

$$\log \phi(t) = \log \left[ 1 + \sum_{i=1}^{\infty} \mu_i t_i \right]$$

Hence

$$\log \phi(t) = t\mu + \log \left[ 1 + (\mu_2 \frac{t^2}{2!} + \mu_3 \frac{t^3}{3!} + \ldots) \right]$$

Since,

$$\log (1+x) = x - 1/2 x^2 + 1/3 x^3 - \ldots,$$

then

$$\log \phi(t) = \mu + (\mu_2 \frac{t^2}{2!} + \mu_3 \frac{t^3}{3!} + \ldots) - \frac{1}{2}(\mu_2 \frac{t^2}{2!} + \mu_3 \frac{t^3}{3!} + \ldots) + \ldots$$

$$= \mu_1 t + \mu_2 \frac{t^2}{2!} + \mu_3 \frac{t^3}{3!} + (\mu_4 - 3\mu_2^2) \frac{t^4}{4!} + \ldots \quad (2)$$
hence, equating coefficients of like powers of \( t \) in (1) and (2) we find:

\[ K_1 = \mu, \quad K_2 = \sigma^2, \quad K_3 = \mu_3, \quad K_4 = \mu_4 - 3\mu_2^2, \quad \text{etc.} \]

The function \( K(t) \) is called the cumulant generating function (c.g.f.).

**Example 4.9.** For the normal distribution

\[
f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty,
\]

\[
\phi(t) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx - \frac{(x-\mu)^2}{2\sigma^2}} \, dx.
\]

Let \( z = \mu + y \), then

\[
\phi(t) = \frac{e^{\mu t}}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ty - \frac{y^2}{2\sigma^2}} \, dy = e^{\mu t} \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \, e^{\frac{t^2\sigma^2}{2}} \, dy
\]

\[
\therefore \phi(t) = e^{\mu t + \frac{t^2\sigma^2}{2}}.
\]

Now, in order to read off the moments we need the expansion of \( \phi(t) \) in series, but this is not very simple. However, the cumulants may be found quite simply, since for this case

\[ K(t) = \log \phi(t) = t\mu + \frac{t^2\sigma^2}{2}. \]

But,

\[ K(t) = K_1 t + K_2 \frac{t^2}{2!} + K_3 \frac{t^3}{3!} + \ldots. \]

hence for the normal distribution

\[ K_1 = \mu, \quad K_2 = \sigma^2, \quad K_3 = 0 \quad \text{for} \quad i > 2. \]

It is important to remember that all cumulants after and including \( K_3 \) for the normal distribution are zero.
Hence, for the normal distribution, it is simpler to read the cumulants, \( K_1 \), for the cumulant generating function, \( K(t) \). If the moments are desired, they may be obtained easily from the cumulants. Hence the use of either the m.g.f. or the c.g.f. depends on the form of the distribution function.

4.8. An inverse problem. It was seen that, if we are given the theoretical distribution, then we may obtain a set of moments \( (\mu, \mu'_2, \mu'_3, \ldots) \). In applied work we may have a large sample of observations and wish to determine from the data some evidence regarding an appropriate theoretical function to represent the assumed parent population distribution.

From the table of values giving the empirical frequency distribution it would be possible to obtain sample moments for the large sample. If it be assumed that these sample moments are "good" estimates of the corresponding moments of some theoretical distribution, then we have the inverse problem of determining uniquely a theoretical distribution having given the moments. This problem is discussed in texts on advanced mathematical statistics and is beyond the scope of this course. The "Pearson system of curves" is an assumed set of continuous functions whose parameters may be expressed in terms of the moments. Hence, estimates of these parameters may be obtained from a large sample and some one of the theoretical curves "fitted" to the empirical frequency distribution; A test of "goodness of fit" may then be affected by the use of Chi-square. The normal distribution is a Pearson type curve.
Closely related to the moment problem mentioned above is the inverse relation of the characteristic function or the m.g.f. to a possible corresponding distribution function. Two theorems from advanced theoretical statistics will be stated without proof and made use of in subsequent derivations.

**Theorem 4.1** - A distribution function is uniquely determined by its characteristic function or, where it exist the moment-generating function.

**Theorem 4.2** - If a distribution function has a characteristic function (m.g.f.) which approaches the characteristic function (m.g.f) of another distribution, the two distributions approach each other.

**Exercises**

**Exercise 4.1** - Find the mathematical expectation for the following distribution:

<table>
<thead>
<tr>
<th>x</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>0.1</td>
<td>0.3</td>
<td>0.5</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**Exercise 4.2** - Given the following probability laws, find $\mu$, $\sigma^2$, $\alpha_3$, and $\alpha_4$ for each:

a. $f(x) = 10x^9$, $0 \leq x \leq 1$

b. $f(x) = x/50$, $0 \leq x \leq 10$.

**Exercise 4.3** - A random variable can assume only two values 1 and 2. Its mathematical expectation is $3/2$. Find $p_1$ and $p_2$.

**Exercise 4.4** - A random variable has the distribution function $f(x) = A + Bx$, $0 \leq x \leq 1$. The mathematical expectation is $1/2$. Find the constants $A$ and $B$. 
Exercise 4.5 - Express $\mu_3$ and $\mu_4$ in terms of moments about zero.

Exercise 4.6 - Use the cumulants for the normal distribution to determine the first four moments about the mean.

Exercise 4.7 - Find $\mu$, $\sigma$, $\alpha_3$, and $\alpha_4$ for the following binomials.

$\left(\frac{1}{6} + \frac{5}{6}\right)^7$, $\left(\frac{1}{3} + \frac{2}{3}\right)^{18}$

Exercise 4.8 - Two other measures of skewness and kurtosis, or departure from normality, are $\gamma_1 = \frac{K_3}{K_2^{3/2}}$, and $\gamma_2 = \frac{K_4}{K_2^2}$. Find $\gamma_1$ and $\gamma_2$ for the normal distribution.

Exercise 4.9 - Determine the first 4 cumulants for the binomial distribution. Verify that $K_{r+1} = pq \frac{dK_r}{dp}$, $r > 1$ for the cumulants obtained.

Exercise 4.10 - Show that $E(x - x)^2$ is a minimum of $x = E(x)$.

Exercise 4.11 - If $x$ has the distribution function $f(x) = 1/2$ on the interval $(0,2)$, find the m.g.f. of $x$ and determine the variance of $x$. This is called a rectangular distribution.

Exercise 4.12 - A penny is tossed 64 times. Find: (a) the expected number of heads, (b) the theoretical standard deviation.

Exercise 4.13 - A pair of dice is thrown 60 times. Find: (a) the expected number of times that the sum 10 appears, (b) the expected value of the square of the standard deviation.

Exercise 4.14 - There are 6 urns containing, respectively: 1 white, 9 black; 2 white, 8 black; 3 white, 7 black; 4 white, 6 black; 5 white, 5 black; 6 white, 6 black balls. One ball is to be drawn from each urn.
(Exercise 4.14) continued -

What is the expected number of white balls taken? Let $x_i$ be a variable which assumes the values 1 or 0 according as to whether the trial results in success or failure. Then, $m = x_1 + x_2 + ... + x_n$ is the number of successes in $n$ trials. But, $E(x_i) = p_i = 1 - (1-p_i) = 0 = p_i$ for $i=1,2,...,n$.

Hence, $E(m) = p_1 + p_2 + ... + p_n$.

Exercise 4.15 - An urn contains $a$ red balls and $b$ black balls, and $c$ balls are drawn. What is the expected number of red balls drawn?

Exercise 4.16 - An urn contains $r$ tickets numbered from 1 to $r$, and $s$ tickets are drawn at a time. What is the expected sum of the numbers on the tickets drawn? Let $x_i$ be the variable attached to the $i^{th}$ ticket which may assume any of the values 1, 2, ..., $r$. Then, $E(x_i) = \frac{1}{r}(1+2+...+r)$. Set $m = x_1+x_1+...+x_r$ and complete the solution by finding $E(m)$.

Exercise 4.17 - Find the m.g.f. for the triangular distribution sketched below:

Exercise 4.18 - Find the m.g.f. for the rectangular distribution $f(x) = 1/w$. Verify that when $w = 1$, the square of the m.g.f. of the Rectangular distribution equals the m.g.f. of the triangular distribution in Exercise 4.17.
Exercise 4.19 - Find $\mu'_2, \mu'_3, \mu'_4$ for the binomial distribution using the formulas for the definitions of these moments. Note: $x^2 = x(x-1) + x$, $x^3 = x(x-1)(x-2) + 3x^2 - 2x$, and $x^4 = x(x-1)(x-2)(x-3) + 6x^3 - 11x^2 + 6x$.

Exercise 4.20 - Find $\alpha_3$ and $\alpha_4$ for the binomial distribution.

Exercise 4.21 - Show that the m.g.f. for the Poisson distribution is $\phi(t) = e^{\mu e^t}$.

Exercise 4.22 - Show that $k_i = \mu(k=1,2,3,\ldots)$ for the Poisson distribution.

Exercise 4.23 - Find $\mu'_1, \mu'_2, \mu'_3, \mu'_4$ for the Poisson distribution:
(a) using the formulas for the definitions of the moments,
(b) using the m.g.f.

Exercise 4.24 - Find $\mu_2, \alpha_3, \alpha_4$ for the Poisson distribution.

Exercise 4.25 - Prove the general formula connecting the moments about $\mu'_1$ with the moments about the origin:

$$
\mu_p = \mu'_p - \mu'_1 \mu'_1 \mu'_{p-1} + \frac{\mu'_1}{21} (\mu'_1)^2 \mu'_{p-2} + \ldots
$$

Use the formula to obtain:

$$
\mu_1 = 0, \mu_2 = \mu^2 - (\mu'_1)^2,
$$

$$
\mu_3 = \mu'_3 - 3\mu'_1 \mu'_2 + 2(\mu'_1)^3 \text{ and}
$$

$$
\mu_4 = \mu'_4 - 4\mu'_1 \mu'_3 + 6(\mu'_1)^2 \mu'_2 - 3(\mu'_1)^4.
$$
Chapter V

Bivariate and Multivariate Distributions and Their Properties

5.1 - Introduction. In the previous chapters single-variate or univariate distributions and their properties have been discussed. It is proposed now to extend the discussion to cases of two or more variates, i.e. bivariate and multivariate distributions. The discussion will apply alike to bivariate and multivariate parent population distributions and bivariate and multivariate derived sampling distributions. The latter distributions will be discussed in Chapter VI.

5.2 - Discrete bivariate distribution. Suppose that for every value of a given variate, \( x \), we also know the values which a second variate, \( y \), can take. Then it will be possible to construct a joint probability distribution, from which can be obtained the probability that any combination of \( x \) and \( y \) will occur in random draws. The bivariate frequency function will be represented symbolically by \( f(x,y) \). The conditions required for a mathematical function \( F(x,y) \) to be used as a cumulative joint probability distribution are analogous to those in Section 3.7 for the univariate case. These conditions imply the following conditions for \( f(x,y) \):

(a) \( f(x,y) \) is non-negative over the \((x,y)\) plane

(b) \( \sum f(x,y) = 1 \), where \( W \) is the entire \((x,y)\) region

(c) \( \sum f(x,y) \) can be computed for any sub-region, \( w \), of \( W \).

Example 5.1. Consider the two-dice problem. Let \( x \) represent the number of spots showing on die 1 and \( y \) the number on die 2 in any one toss of the two dice. The joint probability distribution is given by the following table (p = 1/36):
Table 5.1. Joint Probability Distribution in Two-Dice Problem

<table>
<thead>
<tr>
<th>x</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>p</td>
<td>p</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>p</td>
<td>6p</td>
</tr>
<tr>
<td>2</td>
<td>p</td>
<td>p</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>p</td>
<td>6p</td>
</tr>
<tr>
<td>3</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>4</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>5</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>6</td>
<td>p</td>
<td>p</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>p</td>
<td>6p</td>
</tr>
<tr>
<td></td>
<td>6p</td>
<td>6p</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>6p</td>
<td>36p</td>
</tr>
</tbody>
</table>

The frequency function is \( f(x,y) = 1/36 \) for any pair of values, \((x,y)\), for \( x \) or \( y = 1,2,...,6 \). Note that \( \Sigma f(x,y) = 1 \) over the ranges of both variates, \( 1 \leq x \leq 6, \ 1 \leq y \leq 6 \). The probability that \( x \) and \( y \) lie, at the same time, in ranges \( a \leq x \leq b, \ c \leq y \leq d \), is given by

\[
P(w) = \Sigma f(x,y),
\]

where \( w \) is the sub-region defined by \( a \leq x \leq b, \ c \leq y \leq d \). In the two-dice problem

\[
P(1 \leq x \leq 3, \ 2 \leq y \leq 4) = 9/36 = 1/4.
\]

We may define a cumulative distribution function for the bivariate case in a manner similar to that for \( P(w) \) in (1) above. If \( F(b,d) \) represents the probability that \( x \leq b, \ y \leq d \), then,

\[
P(b,d) = \Sigma f(x,y),
\]

where \( w \) is the sub-region \( (x \leq b, \ y \leq d) \). For the two-dice problem

\[
P(3,2) = 6/36 = 1/6.
\]
If the sub-region \( w \) be allowed to assume all possible values, the \( x_i \) and \( y_j \) will assume all possible pairs of values and (2) in three dimensions becomes analogous to the two dimension step-function of Section 3.3, and may be written as

\[
F(x,y) = \sum_w f(x,y),
\]

over the range of values of the sub-region \( w \). The function defined by (3) is the cumulative bivariate distribution function.

Suppose we extend this process. Add all of the probabilities for a given value of \( x \), say \( x_1 \), then the range of values of the sub-region \( w \) is simply the linear range of \( y \), hence we may write

\[
\sum_y f(x_1,y) = g(x_1),
\]

which is the probability that \( x = x_1 \). If this be thought of as being done for all values of \( x \), then we may write the frequency function of \( x \) as

\[
g(x) = \sum_y f(x,y),
\]

which is called the marginal distribution of \( x \). Similarly, the marginal distribution of \( y \) is

\[
h(y) = \sum_x f(x,y).
\]

As can be seen from the border totals in Table 5.1, the distributions (4) and (5) may be exhibited as in Table 5.2

<table>
<thead>
<tr>
<th>( x = y )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( g(x) = h(y) )</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
<td>1/6</td>
</tr>
</tbody>
</table>

Finally, the marginal distribution and the joint distribution may be used to define the conditional distribution, corresponding to the conditional probability.
discussed in Section 2.5. Using the notation \( f(y|x) \) to mean the probability of \( y \), given \( x \), we know that

\[
f(y|x) = \frac{f(x,y)}{g(x)}, \quad \left[ g(x) \neq 0 \right].
\]

(6)

since

\[
f(x,y) = g(x) f(y|x)
\]

by Law 3 of Section 2.5. The distribution (6) is called the conditional distribution of \( y \). Similarly, the conditional distribution of \( x \) is

\[
f(x|y) = \frac{f(x,y)}{h(y)}, \quad \left[ h(y) \neq 0 \right].
\]

(7)

Now, if \( f(y|x) \) does not depend on \( x \), then \( y \) and \( x \) are said to be independent variates, since

\[
f(x,y) = g(x) \cdot h(y).
\]

(8)

This is true for the two-dice problem since for any \( x = x_1 \), \( f(y|x_1) = 1/6 \).

Hence, by (8),

\[
f(x,y) = 1/6 \cdot 1/6 = 1/36.
\]

for \( x, y = 1, 2, \ldots, 6 \).

5.3 - Continuous bivariate distribution. In the case of two continuous variates, \( x \) and \( y \), the probability that \( x \) is in the interval \((x, x+dx)\) and \( y \) in the interval \((y, y+dy)\) is

\[
f(x,y) dx \, dy.
\]

(9)

The graph of \( z = f(x,y) \) is called the frequency surface. The frequency function \( f(x,y) \) is non-negative, but it may be zero over certain sub-regions of the \((x,y)\) plane.

The probability that \((x,y)\) will fall in some sub-region \( W \) of the \((x,y)\) plane \( W \) is given by

\[
P(W) = \int \int_W f(x,y) dx \, dy,
\]

(10)

since

\[
\int \int_W f(x,y) dx \, dy = 1
\]

(11)
The cumulative distribution function is given by
\[ F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(x,y) \, dy \, dx. \] (12)

The marginal distribution of \( x \) is given by
\[ f(x) = \int_{-\infty}^{\infty} f(x,y) \, dy, \]
and similarly for \( h(y) \).

Finally, the conditional probability that \( y \) lies in the interval \((y, y+dy)\), given that \( x \) is in the interval \((x, x+dx)\) is
\[ f(y|x)\,dy = \frac{f(x,y)\,dx\,dy}{g(x)\,dx} = \frac{f(x,y)\,dy}{g(x)}. \] (13)

Again, if \( f(y|x) \) is independent of \( x \), i.e. if the right side of (13) does not contain \( x \) after algebraic simplification, then \( x \) and \( y \) are said to be independent variates and
\[ f(x,y) = g(x) \cdot h(y). \]

Example 5.2. Given \( f(x,y)\,dx\,dy = e^{-x-y} \, dx \, dy \), \( (x,y \geq 0) \).

(a) \[ F(x,y) = \int_{0}^{x} \int_{0}^{y} e^{-x-y} \, dx \, dy. \] For \( x = x_1 = 1, y = y_1 = 1, \)
\[ F(x_1,y_1) = P(x,y \leq 1) = (1-e^{-x})(1-e^{-y}) = \left(\frac{e-1}{e}\right)^2 = 0.3996. \]

(b) \[ g(x) = \int_{0}^{\infty} e^{-x-y} \, dy = e^{-x}. \]

(c) \[ f(y|x) = \frac{e^{-x-y}}{e^{-x}} = e^{-y}, \] (independent of \( x \)).

Example 5.3. The normal bivariate distribution with means of \( x \) and \( y \) both zero is
\[ f(x,y)\,dx\,dy = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2\rho xy}{\sigma_x \sigma_y} \right] \right\} \, dx \, dy, \]
where $-\infty \leq x \leq \infty$, $-\infty \leq y \leq \infty$.

(a) $g(x) = \int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{a_x \sqrt{2\pi}} e^{-\frac{x^2}{2a_x^2}}$

(b) $f(y|x) = \frac{f(x,y)}{g(x)} = \frac{1}{a_y \sqrt{2\pi} (1-\rho^2)} e^{-\frac{1}{2(1-\rho^2)} \left[ \frac{y-a_y}{\sigma_y} - \frac{a_x}{\sigma_x} \right]^2}$

If $\rho = 0$, $f(y|x) = \frac{1}{a_y \sqrt{2\pi}} e^{-\frac{y^2}{2a_y^2}} = h(y)$;

hence, in this case, $x$ and $y$ are independent. Hence, $\rho$ may be used as a measure of relationship between $x$ and $y$. It is, in fact, the population correlation coefficient between $x$ and $y$.

5.4 - Distributions of functions of discrete variates. In order to obtain certain properties of bivariate and multivariate discrete or continuous distributions, such as expected values, moments, and moment generating functions, it is sometimes necessary to make transformations of the variates so that the summations or integrals may be evaluated. The discrete case will be considered first.

The distribution of a function of $x$, say $z = \psi(x)$, given the distribution of $x$, is simple if there is a one-to-one correspondence between $x$ and $z$; i.e., if for every value of $x$ there is only one value of $z$, and vice versa. In this case, the same probabilities hold for $z$ as for $x$. For example, consider a single die with $f(x) = 1/6$ ($x = 1, 2, \ldots, 6$). Suppose $z = x^2$. In general, there is not a one-to-one correspondence between $x$ and $z$, because $x = \pm z$, resulting in two values of $x$ for each $z$. However, in our die problem, $x$ must be positive; hence $x = +\sqrt{z}$ only. Therefore, $f(z) = 1/6$, for $z = 1, 4, \ldots, 36$. 
On the other hand, suppose \( z = (x-1)(x-2) \), or \( x = \frac{3 + \sqrt{1+4z}}{2} \). Then, even for \( x \) always positive, there is not a one-to-one correspondence between \( x \) and \( z \), since, say, for \( z = 0 \), \( x = 1 \) or 2. Hence, in this case,

\[
\begin{align*}
1/3 & \quad \text{for } z = 0, \ i.e. \ two \ values \ of \ 0 < x \leq 6, \\
f(z) & = 1/6 \quad \text{for } z = 2, 6, 12, 20, \ one \ integral \ value \ of \ 0 < x \leq 6, \\
0 & \quad \text{elsewhere, no integral value of } 0 < x \leq 6.
\end{align*}
\]

Again, if \( z = (x-1)(x-2) \ldots (x-6) \)

\[
\begin{align*}
1 & \quad \text{for } z = 0, \ i.e. \ six \ values \ of \ 0 < x \leq 6 \\
f(z) & = 0 \quad \text{elsewhere, no values of } 0 < x \leq 6.
\end{align*}
\]

If we consider a bivariate distribution, such as that of the two-dice problem the distribution of a function of the two variates is still simple.

For example, consider the distribution of \( f_2(w) \), where \( w = xy \), then

\[
\begin{align*}
 f_2(1) & = f(1,1) = 1/36 \\
f_2(2) & = f(2,1) + f(1,2) = 2/36 \\
f_2(3) & = f(3,1) + f(1,3) = 2/36 \\
f_2(4) & = f(4,1) + f(2,2) + f(1,4) = 3/36 \\
\vdots \\
f_2(7) & = 0 \\
\vdots \\
f_2(36) & = f(6,6) = 1/36
\end{align*}
\]

Here again \( \sum f_2(w) = 1 \) over the possible range.

5.5 - Distributions of functions of continuous variates. The distribution of \( z \) is \( f(x)dx \), \( x \) defined in the range \( x_1 \) to \( x_2 \), we seek the distribution of \( z = g(x) \). If there is a one-to-one correspondence between \( x \) and \( z \), and \( x \) can be solved uniquely in terms of \( z \), then \( x = \psi(z), \ dx = \psi'(z)dz, f(z) = \int \psi'(z) \)

\[
\]
and the limits are \( z_1 = x(z_1), z_2 = x(z_2) \). The probability distribution of 
\( z \), with these conditions, will be \( f(z) \psi(x) \psi'(x)dz \) over the range \( z_1 \) to \( z_2 \).

**Example 5.4.** If \( f(x)dx = 2(1-x)dx \) for \( 0 \leq x \leq 1 \), we find the distribution of 
\( z \), where \( z = x^2 \), as follows:

Since \( z \) is always positive, there is a one-to-one correspondence between 
\( z \) and \( x \), i.e. \( x = \sqrt{z} \), hence

\[
I(x) = 2(1 - \sqrt{z}) \frac{1}{2} \sqrt{\frac{1}{z}} dx
\]

\[
= \left( \frac{1}{2} - 1 \right) dx,
\]

\( 0 \leq x \leq 1 \).

**Example 5.5.** If \( f(x)dx = \frac{1-x}{2} dx \), \( -1 \leq x \leq 1 \), then \( x = \sqrt{z} \) for \( z = x^2 \). For
positive values of \( z \), \( x = \sqrt{z} \) and for negative values of \( z \), \( x = -\sqrt{z} \). In
this case

\[
I(z) = \begin{cases} 
\frac{1}{2} (z - \frac{1}{2}) dx & x \geq 0 \\
\frac{1}{2} (z + \frac{1}{2}) dx & x \leq 0, 0 \leq z \leq 1
\end{cases}
\]

If we wish, we may add the two functions to obtain a single function

\[
I(z) = \frac{1}{2} z - \frac{1}{2},
\]

\( 0 \leq z \leq 1 \),

but this is not possible in all cases.

The distribution of a function of two continuous variates, \( x \) and \( y \) is
more complex mathematically. We wish to derive the simultaneous distribution
of \( u \) and \( v \), \( f(u,v)du dv \), where \( u = u(x,y) \) and \( v = v(x,y) \). In case we wish
only the distribution of \( u \), then \( y \) is integrated out between its limits of
integration leaving some \( f(u)du \). In such cases it is necessary in general
to assume a one-to-one correspondence between \((x,y)\) and \((u,v)\). It will then
be possible to make the inverse solution:

\[
x = x(u,v) \text{ and } y = y(u,v).
\]
The probability distribution of \( f(x,y) \, dx \, dy \) then becomes

\[
f \left[ x(u,v), y(u,v) \right] \left| J \right| \, du \, dv,
\]

where \( J \) is the Jacobian of the transformation, i.e.

\[
J = \begin{vmatrix}
\frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\
\frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}
\end{vmatrix}
\text{ or } \frac{1}{J} = \begin{vmatrix}
\frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\
\frac{\partial v}{\partial x} & \frac{\partial v}{\partial y}
\end{vmatrix}
\]

This implied, of course, that these partial derivatives exist. If the second form is used, then \( J \) must be evaluated at \( x = x(u,v), y = y(u,v) \).

The limits for \( u \) and \( v \) must be determined individually for each problem.

The limits for the first variable in the integral may be functions of the second variable.

**Example 5.6.** Let us find \( f(u,v) \, du \, dv \), where \( f(x,y) \, dx \, dy = e^{-x-y} \, dx \, dy \),

\[
0 \leq x \leq \infty \quad 0 \leq y \leq \infty, \quad u = x+y, \quad v = xy.
\]

Then,

\[
x = \frac{uv}{1+v}, \quad y = \frac{u}{1+v},
\]

and

\[
J = \begin{vmatrix}
\frac{v}{1+v} & \frac{u}{(1+v)^2} \\
\frac{1}{1+v} & \frac{-u}{(1+v)^2}
\end{vmatrix} = \frac{-u}{(1+v)^2}
\]

or

\[
\frac{1}{J} = \begin{vmatrix}
1 & 1 \\
\frac{1}{y} & -\frac{x}{y^2}
\end{vmatrix} = \frac{(x+y)}{y} = \frac{(1+v)^2}{u}
\]

\[
\therefore \ f(u,v) \, du \, dv = e^{-u} \frac{u}{(1+v)^2} \, du \, dv.
\]

Since, if \( u \) is fixed, \( v \) can assume any value, then the limits of integration for \( u \) and \( v \) are \( 0 \leq u \leq \infty, \ 0 \leq v \leq \infty \).
To find the distribution of \( u \) we have

\[
f(u)du = \left[ \int_{0}^{\infty} \frac{1}{(1+y)^{w}} dv \right] du \]

\[= u e^{-u} du, \quad 0 \leq u \leq \infty\]

Similarly,

\[
f(v)dv = \frac{1}{(1+v)^{w}} dv, \quad 0 \leq v \leq \infty\]

Note that \( u \) and \( y \) are independent in this case.

5.6. **Expected values for bivariate distributions.** For any bivariate frequency function, \( f(x,y) \), the expected value of any function of \( x \) and \( y \), say, \( \psi(x,y) \) is

\[
E \left[ \psi(x,y) \right] = \sum_{x,y} \psi(x,y) f(x,y) dx \ dx dy,
\]

where \( \sum \) is the entire region of \( (x,y) \). The following specializations of \( \psi(x,y) \) will enable us to derive some simple rules for operating with expected values. Continuous variates will be used in the derivations, but similar rules will hold for discrete variates.

1. \( E(c) = c \), where \( c \) is a constant.
2. \( E(cx) = c \int_{-\infty}^{\infty} x \left[ \int_{-\infty}^{\infty} f(x,y) dy \right] dx = c \int_{-\infty}^{\infty} x g(x) dx = c E(x) \)
3. \( E(x+y) = \int_{-\infty}^{\infty} x \left[ \int_{-\infty}^{\infty} f(x,y) dy \right] dx + \int_{-\infty}^{\infty} y \left[ \int_{-\infty}^{\infty} f(x,y) dx \right] dy \)

\[= \int_{-\infty}^{\infty} x g(x) dx + \int_{-\infty}^{\infty} y h(y) dy \]

\[= E(x) + E(y) \]
4. \( E(xy) = \int_{-\infty}^{\infty} x \left[ \int_{-\infty}^{\infty} y f(x,y) dy \right] dx = \int_{-\infty}^{\infty} x h_{1}(x) dx, \)

where

\[h_{1}(x) = \int_{-\infty}^{\infty} y f(x,y) dy.\]
Note that \( E(xy) \) can be evaluated, if the integrations can be performed, even though \( x \) and \( y \) are not independent.

(5) If \( x \) and \( y \) are independent, then \( f(x,y) = g(x)h(y) \) and hence

\[
E(xy) = \int_{-\infty}^{\infty} x g(x) dx \int_{-\infty}^{\infty} y h(y) dy = E(x)E(y).
\]

5.7 - Moments. The \( r \)-th moment of \( x \) and the \( s \)-th moment of \( y \) about the origin is given by

\[
\mu_{rs} = E(x^r y^s) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^r y^s f(x,y) dx \, dy.
\]

Let the mean of \( x \) be \( \mu_{10} \) and the mean of \( y \) be \( \mu_{01} \), then

\[
\mu_{rs} = E[(x-\mu_{10})^r(y-\mu_{01})^s] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-\mu_{10})^r(y-\mu_{01})^s f(x,y) dx \, dy.
\]

Example 5.6. To find \( c_x^2 \), defined as equal to \( \mu_{20} \), we have

\[
\mu_{20} = E(x-\mu_{10})^2 = \mu_{20} - (\mu_{10})^2.
\]

\[
\therefore \quad c_x^2 = \mu_{20} - (\mu_{10})^2.
\]

Similarly,

\[
c_y^2 = \mu_{20} - (\mu_{01})^2.
\]

Example 5.7. To find \( c_{xy} \), defined as equal to \( \mu_{11} \) and called the covariance of \( x \) and \( y \), we have

\[
\mu_{11} = E[(x-\mu_{10})(y-\mu_{01})] = E(xy) - \mu_{10} \mu_{01}.
\]

\[
\therefore \quad c_{xy} = E(xy) - \mu_{10} \mu_{01}.
\]

If \( x \) and \( y \) are independent, \( E(xy) = \mu_{10} \mu_{01} \), hence \( c_{xy} = 0 \) for this condition. The correlation, \( \rho_{xy} \), between \( x \) and \( y \) is defined as the non-dimensional quantity

\[
\rho_{xy} = \frac{c_{xy}}{c_x c_y}, \quad -1 \leq \rho \leq 1.
\]

Hence, if \( c_{xy} = 0 \), then \( \rho_{xy} = 0 \).
Example 5.6. To find the variance of \((x+y)\) we have
\[
\sigma_{x+y}^2 = \mathbb{E}\left[(x-\mu_x + y - \mu_y)^2\right] = \sigma_x^2 + \sigma_y^2 + 2 \rho \sigma_x \sigma_y \\
= \sigma_x^2 + \sigma_y^2 + 2 \rho \sigma_x \sigma_y.
\]
If \(x\) and \(y\) are independent, i.e., \(\sigma_{xy} = 0\), then \(\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2\).

Example 5.8. To find \(\mathbb{E}(xy)\) for the bivariate normal distribution, with \(\mu_{10} = 0, \mu_{01} = 0\), we have
\[
\mathbb{E}(xy) = k \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \, e^{-\phi} \, dx \, dy,
\]
where
\[
k = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1-\rho^2}}
\]
and
\[
\phi = \frac{1}{2(1-\rho^2)} \left[ \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \rho \frac{xy}{\sigma_x \sigma_y} \right].
\]
Note that for the bivariate normal \(\rho_{xy}\) has been abbreviated to \(\rho\).

The function \(\phi\) may be written
\[
\phi = \frac{1}{2(1-\rho^2)} \left[ \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \rho \frac{xy}{\sigma_x \sigma_y} \right] + \frac{y^2}{2\sigma_y^2}
\]
\[
= \frac{1}{2} \left[ \frac{x^2}{1-\rho^2} + \frac{y^2}{\sigma_y^2} \right], \text{ where } t = \frac{x}{\sigma_x} - \frac{\rho y}{\sigma_y}.
\]

Then, using the methods of Section 5.5
\[
\mathbb{E}(xy) = k \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\sigma_x^2(t \sigma_y + \rho y)}{\sigma_y^2} \, e^{-t^2/(2(1-\rho^2))} \, dt \, dy
\]
\[
= \frac{\rho \sigma_x^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{y^2}{\sigma_y^2} \, e^{-y^2/(2\sigma_y^2)} \, dy
\]
\[
= \rho \sigma_x \sigma_y.
\]

Example 5.9. Consider
\[
f(x,y)dx \, dy = e^{-x-y} \, dx \, dy, \quad 0 \leq x \leq \infty, \quad 0 \leq y \leq \infty.
\]
Then,

\[ E(x) = \mu_{10} = \int_0^\infty e^{-y} \left[ \int_0^\infty xe^{-x}dx \right] dy = 1, \]

\[ E(y) = \mu_{01} = 1, \]

\[ \mu_{20} = \mu_{02} = 2, \]

\[ \sigma_x^2 = \sigma_y^2 = 1, \]

\[ \mu_{11} = \int_0^\infty x e^{-x}dx \int_0^\infty y e^{-y}dy = 1, \]

\[ \sigma_{xy} = \mu_{11} - \mu_{10} \mu_{01} = 0, \]

and

\[ \rho_{xy} = 0. \]

Example 5.10. Consider the following discrete bivariate distribution with \( p = 1/12 \):

<table>
<thead>
<tr>
<th>x</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2p</td>
<td>2p</td>
<td>2p</td>
<td>6p</td>
</tr>
<tr>
<td>1</td>
<td>p</td>
<td>4p</td>
<td>p</td>
<td>6p</td>
</tr>
<tr>
<td>Total</td>
<td>3p</td>
<td>6p</td>
<td>3p</td>
<td>12p</td>
</tr>
</tbody>
</table>

Then,

\[ \mu_{10} = \sum \sum x f(x,y) = 0.5; \mu_{01} = \sum \sum y f(x,y) = 1.0; \]

\[ \mu_{20} = \sum \sum x^2 f(x,y) = 0.5; \mu_{02} = \sum \sum y^2 f(x,y) = 1.5; \]

\[ \sigma_x^2 = 0.5 - 0.25 = 0.25; \sigma_y^2 = 1.5 - 1.0 = 0.5; \]

\[ \mu_{11} = \sum \sum xy f(x,y) = 0.5; \sigma_{xy} = 0.5 - 0.5 = 0; \]

and \( \rho = 0. \)

Note, that we cannot state that \( x \) and \( y \) are independent in this case, even though \( \rho = 0 \), since \( f(x|y) \) is not the same for all values of \( y \). It can be seen that \( f(x|y = 0) = 2/3, 1/3, f(x|y = 1) = 1/3, 2/3, \) and \( f(x|y = 2) = 2/3, 1/3. \)
5.8 - Moment and cumulant generating functions. For the bivariate case, the moment-generating function about \( \mu_{10} \) and \( \mu_{01} \) is defined as:

\[
\phi(t_x, t_y) = E \left\{ e^{T} \right\} = \iint e^{T(x,y)} dx \, dy,
\]

where \( T = (x - \mu_{10})t_x + (y - \mu_{01})t_y \), and the double integration is performed over the ranges of \( x \) and \( y \).

Then,

\[
\phi(t_x, t_y) = \sum_{r,s=0}^{\infty} \mu_{rs} \frac{t_x^r t_y^s}{r!s!},
\]

and

\[
\mu_{rs} = \frac{\partial^{r+s} \phi(t_x, t_y)}{\partial t_x^r \partial t_y^s} \bigg|_{t_x = t_y = 0}.
\]

Note that \( \mu_{00} = 1, \mu_{01} = \mu_{10} = 0 \).

The cumulant-generating function in this case is given by

\[
K = \log \phi(t_x, t_y) = \log \phi + \mu_{10} t_x + \mu_{01} t_y + \ldots
\]

\[
= \sum_{r,s=0}^{\infty} k_{rs} \frac{t_x^r t_y^s}{r!s!}
\]

and

\[
k_{rs} = \frac{\partial^{r+s} K}{\partial t_x^r \partial t_y^s} \bigg|_{t_x = t_y = 0}.
\]

If \( f(x,y) = g(x)h(y) \), the moments of \( x \) and \( y \) may be computed separately.

In this case

\[
\phi(t_x, t_y) = \int e^{T_x} g(x) \, dx \cdot \int e^{T_y} h(y) \, dy = \phi(t_x) \phi(t_y),
\]

where

\[
T_x = (x - \mu_{10})t_x, \quad T_y = (y - \mu_{01})t_y,
\]

and the two integrations are taken over the respective ranges of \( x \) and \( y \).

Similarly, for this case,

\[
K = \log \phi'(t_x) + \log \phi'(t_y) = K(t_x) + K(t_y).
\]

5.9 - Extension to \( k \)-variates. Let

\[
f \left\{ x_1 \right\} \prod_{i=1}^{k} f(x_i)
\]
represent the joint probability distribution of the \( k \) variates, \( x_1, x_2, \ldots, x_k \), where \( f\{x_i\} \) is the frequency function of the \( k \) variates and \( \pi dx_1 = dx_1 dx_2 \ldots dx_k \). The following properties are those for the continuous case, but they can be applied equally well to discrete variates.

(a) \( F(\omega) = \int_{\omega} \ldots \int_{\omega} f\{x_1\} \prod_{i=1}^{k} dx_i \),

where \( \omega \) is a sub-region in the \( k \)-dimensional space.

(b) \( g(x_1) = \int_{\omega} \ldots \int_{\omega} f\{x_1\} \prod_{i=2}^{k} dx_i \)

(c) \( f(x_1|z_2,\ldots,z_k) = \frac{f\{x_1\}}{f(x_2,\ldots,x_k)} \),

where

\[
f(x_2,\ldots,x_k) = \int f\{x_1\} dx_1
\]

over the range of \( x_1 \).

(d) For a transformation \( u_i = u_i\{z_i\}, \quad i = 1,2,\ldots,k \)

\[
\begin{vmatrix}
\frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} & \ldots & \frac{\partial u_1}{\partial x_k} \\
\frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} & \ldots & \frac{\partial u_2}{\partial x_k} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial u_k}{\partial x_1} & \frac{\partial u_k}{\partial x_2} & \ldots & \frac{\partial u_k}{\partial x_k}
\end{vmatrix}
\]

\[
(\begin{array}{c}
e\{z_1\} = \int_{\omega} \ldots \int_{\omega} e^{\{z_1\}} f\{x_1\} \prod_{i=1}^{k} dx_i.
\end{array})
\]

Example 5.11. The multinomial distribution is a generalization of the binomial distribution. Any one of the events \( y_1, y_2, \ldots, y_k \) can occur with respective probabilities \( p_1, p_2, \ldots, p_k \) on a single trial, \( \sum_{i=1}^{k} p_i = 1 \). If \( n \) trials are made, the probability that \( y_1 \) occurs \( x_1 \) times, \( y_2 \) occurs \( x_2 \) times, etc. \( \sum_{i=1}^{k} x_i = n \).
\[ f(x_i) = \frac{n!}{k} \frac{k}{\pi} x_i^{k-1} \]

This is the general term of the expansion of \((p_1 + p_2 + \ldots + p_k)^n\).

For example, a single die can show the number 1, 2, \ldots, 6 on the upper face with equal probabilities, \(p_i = 1/6\).

For \(f(x_i)\) the m.g.f. is

\[ \phi'\{\theta_i\} = E \left[ e^{k \theta_i x_i} \right] = \sum_{x_i=1}^{\infty} \frac{n!}{x_i!} e^{k \theta_i x_i} \]

\[ = \sum_{x_i=1}^{\infty} \frac{n!}{x_i!} \pi(p_i e^{\theta_i}) \]

Then

\[ \mu_{ii} = \left( \frac{\partial}{\partial \theta_i} \right) f(\theta_i) = 0 = np_i \]

\[ \mu_{ii} = \left( \frac{\partial^2}{\partial \theta_i^2} \right) f(\theta_i) = 0 = np_i + n(n-1)p_i^2 \]

and

\[ \sigma^2 = \mu_{ii} - (\mu_{ii})^2 = np_i - np_i^2 = np_i(1-p_i). \]

Also,

\[ \sigma_{ij} = \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} \right) f(\theta_i) = 0 = -\mu_i \mu_j = -np_i p_j. \]
Exercises

Exercise 5.1 - The marginal distribution of $x$ for a bivariate distribution function $f(x,y)$ is

$$g(x) = \int f(x,y)dy.$$  

The limits for $y$ must be determined by the region $W$ within which $f(x,y)$ is defined. If $W$ includes the entire $(x,y)$ plane, the limits are $(-\infty, \infty)$. However, consider a problem of this nature: $f(x,y) = kxy$ for $0 \leq y \leq 1$, then

$$g(x) = \int_{x}^{1} kxy \, dy \quad 0 \leq x \leq 1$$

and

$$h(y) = \int_{0}^{y} kxy \, dx \quad 0 \leq y \leq 1.$$  

Find the explicit values of $y(x)$ and $h(y)$.

Exercise 5.2 - Given the bivariate frequency function $f(x,y) = 2/a^2$, $0 \leq x \leq y$, $0 \leq y \leq a$. Show that: (a) $\int_{x}^{y} f(x,y) \, dx \, dy = 1$, over the respective ranges of $x$ and $y$. (b) $y(x) = 2/a^2 \, (a-x)$, (c) $h(y) = \frac{y}{2}$, (d) $f(x|y) = 1/y$, (e) $\mu_{x0} = \frac{a}{3}$, $\mu_{0x} = \frac{2a}{3}$, (f) $\rho = \frac{1}{2}$, (g) $\mu_{yx} = \frac{ax}{2}$, (h) $\mu_{xy} = y/2$.

Exercise 5.3 - Given a continuous bivariate frequency function $f(x,y)$ and the corresponding marginal distributions, $g(x)$ and $h(y)$. Set up the integrals for the following: (a) Mean and variance of $y$ for a given $x(\mu_{yx}, \sigma^2_{yx})$. (b) Mean and variance of $x$ for a given $y$.

Exercise 5.4 - In Exercise 5.3 (a) the mean value of $y$ for a given $x(\mu_{yx})$ is a function of $x$ and hence defines a curve in the $(x,y)$ plane called the curve of regression of $y$ on $x$. If the regression of $y$ on $x$ is linear ($\mu_{yx} = \alpha + \beta x$) then

$$\mu_{yx} = \int_{\infty}^{-\infty} y f(x,y) \, dy = \alpha + \beta x$$

or

$$\int_{-\infty}^{\infty} y f(x,y) \, dy = \alpha y(x) + \beta x y(x).$$
By integrating each side of the last equation with respect to \( x \), show that
\[
\mu_{10} = \mu_{1} + \beta \mu_{10}.
\]

Before integrating, as above, multiply both sides by \( x \) and show that
\[
\mu_{11} = \alpha \mu_{10} + \beta \mu_{20}.
\]

Hence, find the values of \( \alpha \) and \( \beta \) in terms of the moments of the original distribution, and show that
\[
\mu_{y|x} = \mu_{10} + \rho \frac{\partial \mu_{1} }{\partial x} (x - \mu_{10}).
\]

**Exercise 5.5** - In Exercise 5.3 (a) if the variance of \( y \) for a given \( x \) is averaged over all values of \( x \) we have
\[
\sigma_{y|x}^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sigma(x) \left[ (y - \mu_{y|x})^2 \frac{f(x,y)}{g(x)} \right] dy \, dx
\]
\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y - \mu_{y|x})^2 f(x,y) \, dy \, dx.
\]

If \( \mu_{y|x} = \alpha + \beta x \), as given in Exercise 5.4, show that
\[
\sigma_{y|x}^2 = \sigma_y^2 (1 - \rho^2).
\]

**Exercise 5.6** - Work Exercises 5.3 a, 5.4, and 5.5 for a normal bivariate distribution.

**Exercise 5.7** - As an example of the distribution of a function of two variables for a discrete bivariate distribution consider the distribution of the sum of two independent Poisson variates:
\[
f(x,y) = e^{-m_1} \frac{m_1^x}{x!} e^{-m_2} \frac{m_2^y}{y!}.
\]

Let \( z = x+y \), or \( x = z-y \), then
\[
f(z,y) = e^{-(m_1+m_2)} \frac{m_1^{(z-y)} m_2^y}{(z-y)! \, y!}.
\]

Sum out the variable \( y \) over the range \( 0 \leq y \leq z \) and show that
\[
f(z) = e^{-(m_1+m_2)} \frac{(m_1+m_2)^z}{z!}.
\]
Hence, show that the sum of two independent Poisson variates is distributed as a single Poisson with a mean equal to the sum of the two single means \( m = m_1 + m_2 \).

Exercise 5.8 - Given \( f(x, y) = 6(1-x-y) \) for \((x,y)\) contained within the triangle bounded by \( x = 0, \ y = 0, \ x+y = 1 \).

(a) Find the means and variances of \( x \) and \( y \) and the covariance of \( x \) and \( y \).

(b) Find the equation of the regression line of \( y \) on \( x \) and \( \sigma^2_{y|x} \).

Exercise 5.9 - Given \( f(x, y) = kx^y(1-x-y) \) over the same triangle as in Exercise 5.8.

(a) Find the value of \( k \) which makes \( f(x, y) \) a frequency distribution.

(b) Find the marginal distribution \( y(x) \).

(c) Find \( \mu_y \).

Exercise 5.10 - If \( x \) is a discrete variate having the Poisson distribution

\[
g(x) = \frac{e^{-m} m^x}{x!}
\]

and \( y \) is another discrete variate having the binomial distribution

\[
f(y|x) = p^y q^{x-y}
\]

show that the marginal distribution of \( y \) is

\[
\frac{(mp)^y \ e^{-mp}}{y!}
\]

Exercise 5.11 - Given that \( x \) and \( y \) are normally distributed with zero means and variances \( \sigma_x^2 \) and \( \sigma_y^2 \). Find the distribution of \( z = x + y \).

Exercise 5.12 - If \( x \) and \( y \) represent the number of dots appearing on dice A and B, respectively, what is the probability that in throwing the two dice \( x + 2y \leq 6 \)?
Exercise 5.13 - Given \( f(x|y) = \frac{y^x e^{-x}}{x!} \) and \( h(y) = e^{-y} \), where \( x \) is discrete \((x = 0, 1, \ldots)\) and \( y \) continuous \((y \geq 0)\). Show that \( y(x) = (1/2)^{x+1} \).

Exercise 5.14 - Given two continuous variates \( x \) and \( y \) \((x, y \geq 0)\) with the frequency distribution \( f(x,y) = 2(1+x+y)^3 \). Find \( f(u) \) and \( P(u \leq 1) \), where \( u = xy \).

Exercise 5.15 - Given \( f(x,y) = 1 \) over the square \((0 \leq x, y \leq 1)\). Show that \( P(xy > u) = 1 - u + u \log u \).
Chapter VI
Derived Sampling Distributions and Orthogonal Linear Functions

6.1 - Introduction. For the sequence of the statistical method - specification, distribution, estimation, and tests of hypotheses - we have considered certain parent populations and their properties which are usually specified in applied statistical investigations. After problems of specification it would seem logical to discuss problems of estimation next in order. Such problems involve determining what functions of the sample observations should be used to "best" estimate the parameters of the specified population distribution, where "best" must be defined in some exact manner. A sample estimate of a population parameter is called a statistic. For example, if \( x_1, x_2, \ldots, x_n \) is a sample of \( n \) observations, specified as having been drawn in some manner from a normal distribution with mean \( \mu \), and variance \( \sigma^2 \), what two functions of the observations (statistics) should be used to "best" estimate \( \mu \) and \( \sigma^2 \)? It seemed appropriate to earlier workers in statistics to use the sample mean, \( \bar{x} = \frac{\sum x}{n} \) and the sample variance, \( (s^1)^2 = \frac{\sum (x - \bar{x})^2}{n} \), as sample estimates of the two populations parameters \( \mu \) and \( \sigma^2 \). However, as will be shown later, in cases where the sample is a random sample, the "best" sample estimate of \( \sigma^2 \) is \( s^2 = \frac{\sum (x - \bar{x})^2}{n - 1} \). A random sample is a sample drawn in such a manner that the probability of obtaining any member is independent of the probability of obtaining any other member.

A less restrictive method of estimating the value of a parameter, \( \theta \), is a method which derives limits \( c_1 \) and \( c_2 \) which are functions of the sample values \( (x_1, x_2, \ldots, x_n) \). The interval \( (c_1, c_2) \) is called the confidence interval. This confidence interval is determined so that in repeated sampling from the same population, the interval \( (c_1, c_2) \) will contain the parameter \( \theta \) a certain percentage of the time. The limits are thus functions of the sample and this
percentage, which is called the confidence probability. It is understood that in each repeated sample a new set of confidence limits is determined. The concepts of confidence limits were introduced by Neyman (1).

Various methods of estimating the confidence interval shall be discussed in subsequent sections of this chapter and in later chapters. Fisher (2) uses the term fiducial limits to indicate essentially the same concept.

In the chronological development of modern statistical methodology, however, the distributions of statistics and the distributions of certain functions of these statistics used in making tests of hypotheses and setting confidence limits did not await upon the development of a sound theory of estimation. In this chapter then we shall assume, for the time being, that certain statistics are the "best" estimates of the corresponding population parameters and that certain functions of these statistics used in making tests of hypotheses and setting confidence limits are the "best" such functions.

6.2 - Derived sampling distribution problems. In a typical applied statistical investigation, i.e. a sample survey or an experiment, we specify that the observations obtained by some sampling process are drawn from some particular parent population distribution. We then calculate certain functions of the observations (statistics) as estimates of the parameters of the specified population. Now, in order to set confidence limits for the population parameters or to make tests of hypotheses concerning the population parameters it is necessary to know the probability distributions of the statistics, e.g. of $\bar{x}$ and $s^2$ for a normal parent population, and of certain functions of the statistics, e.g. of $\chi^2$, $t$, and $F$. Mathematically these probability distributions are derived from the specified parent population distributions and hence are called derived sampling distribution.
6.3 - Random and systematic samples. In applied statistics two kinds of samples are in common use: (a) the random sample as defined above, and (b) the systematic sample. In the latter the first member may be chosen at random but subsequent members depend upon the position or value of the preceding members or all members of the sample may be chosen systematically including the first member.

Soil sampling - to study nutritive (3) and other components (4) of the soil - in most cases makes use of some form of systematic sampling. Soil samples are selected from various spots in a field, and chemical determinations are made to determine what nutrients are required to bring the soil up to a suitable productive capacity. It would be possible to lay a grid down on the field and select the samples at random from the grid. This method, however, presents many practical difficulties such as the exact determination of the selected sample point and the excessive time consumed in finding these points, especially if the field is irregular in both boundary and contour. It is much easier to select the first point at random by selecting two random numbers to determine the two coordinates of the starting point and then proceed a certain number of paces from this point to the next one by a predetermined route.

Another systematic method is to predetermine a definite route and collect samples only along this path. This latter method is used quite extensively in sampling forests stands (5) to determine the amount of usable timber.

Most economic data are far from random, but have been analyzed in many instances in the past as though they were random due to lack of techniques for analyzing non-random data. More exact methods of analyzing this type of non-random data have been devised. The chief obstacle to the use of the more exact techniques is computational. The rapid development of electronic computers may overcome this difficulty.
The derived sampling distributions obtained in the next chapter will assume a random sample of \( n \) observations unless otherwise stated. Randomness in the sample must be insured by the use of an objective method of selection. Tables of random numbers provide such a means, and have been made available by Tippett (6), Fisher and Yates (7), Snedecor (6) and others.

6.4 - Linear functions. In the derivations of sampling distributions in the next chapter it will frequently be necessary to know the distribution of some linear function of the members of a sample. Let such a linear function be given by

\[
\mathcal{L} = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n = \sum_{i=1}^{n} a_i x_i,
\]

where \( a_i \) is some fixed constant and the sample, here not necessarily random, is represented by \( x_1, x_2, \ldots, x_n \) or \( \{x_i\} \).

6.5 - Properties of linear functions. It follows that

\[
E(\mathcal{L}) = \sum_{i=1}^{n} a_i E(x_i) = \sum_{i=1}^{n} a_i \mu_i
\]

Also,

\[
\sigma_\mathcal{L}^2 = E\left[ (\mathcal{L} - E(\mathcal{L}))^2 \right] = E\left[ \sum_{i=1}^{n} a_i (x_i - \mu_i)^2 \right] = \sum_{i=1}^{n} a_i^2 \sigma_i^2 + 2 \sum_{i=1}^{n} a_i a_j \sigma_{ij}.
\]

If \( \{x_i\} \) is a random sample, then \( \sigma_{ij} = 0 \), and

\[
\sigma_\mathcal{L}^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2,
\]

but this will not be true for a systematic sample since \( \sigma_{ij} \) will not be zero in this case.

Finally, if all \( \mu_i = \mu_1 \) and all \( \sigma_i^2 = \sigma^2 \), then

\[
E(\mathcal{L}) = \mu_1 \sum_{i=1}^{n} a_i.
\]
and
\[ c_1^2 = \sigma^2 \sum_{i=1}^{n} x_i^2. \]

**Example 6.1.** If each \( a_i = 1/n \), then \( \mathcal{L} = \sum_{i=1}^{n} x_i \) which is the sample mean, \( \bar{x} \).

Further, for a random sample
\[ E(\bar{x}) = \frac{1}{n} (n \mu_1) = \mu_1. \]

Also, if the sample be random,
\[ \frac{\sigma^2}{\bar{x}} = E((\bar{x} - \mu_1)^2) = \sigma^2 \frac{n}{\bar{x}} = \frac{1}{n^2} \frac{\sigma^2}{\bar{x}}. \]

In Section 6.1 mention was made of a "best" sample estimate of a population parameter. One property usually desired in a "best" estimate is that of being unbiased. An estimate of a population parameter is said to be unbiased if its expected value is equal to the population parameter.

It follows for Example 6.1 that \( \bar{x} \) is an unbiased estimate of \( \mu_1 \) whether the sample is random or systematic. However, using an analogous method to that for a random sample for obtaining an estimate of the variance of a systematic sample does not lead to an unbiased estimate.

**Example 6.2.** Let \( \mathcal{L} = \sum_{i=1}^{n} a_i (x_i - \mu_1) \), for a random sample from a population with \( \mu_{1i} = \mu_1 \), \( c_1^2 = \sigma^2 \), then
\[ E(\mathcal{L}) = \sum_{i=1}^{n} a_i E(x_i - \mu_1) = 0. \]

Also,
\[ c_1^2 = E\left[ \mathcal{L} - E(\mathcal{L}) \right]^2 = E\left[ \sum_{i=1}^{n} a_i^2 (x_i - \mu_1)^2 + 2 \sum_{i \neq j} a_i a_j (x_i - \mu_1)(x_j - \mu_1) \right] \]
\[ = \sigma^2 \sum_{i=1}^{n} a_i^2. \]
Example 6.3. To find $\psi(S)$ where $S = \sum_{i=1}^{n} (x_i - \bar{x})^2$, we find

$$S = \sum_{i=1}^{n} \left[ (x_i - \mu_i) - (\bar{x} - \mu_i) \right]^2 = \sum_{i=1}^{n} (x_i - \mu_i)^2 - n(\bar{x} - \mu_i)^2.$$ 

But,

$$\sum_{i=1}^{n} (x_i - \mu_i)^2 = n \sigma^2.$$ 

Also,

$$\mathbb{E}(\bar{x} - \mu_i)^2 = \frac{\sigma^2}{n} \text{ by Example 6.1.}$$ 

Hence,

$$\mathbb{E}(S) = n \sigma^2 - \sigma^2 = (n-1)\sigma^2.$$ 

Now, if we set $s^2 = \frac{S}{n-1}$, then $\mathbb{E}(s^2) = \sigma^2$, and therefore $s^2$ is an unbiased estimate of $\sigma^2$ for a random sample.

For a discussion of the above problems for systematic samples consult W.G. Madow and Lillian Madow (9), Lillian Madow (10) and Cochran (11).

6.6 - Orthogonal linear forms. Consider the two linear forms $l_1 = \sum_{i=1}^{n} a_i x_i$ and $l_2 = \sum_{i=1}^{n} b_i x_i$. Since, $\mathbb{E}(l_1) = \sum_{i=1}^{n} a_i \mu_i$ and $\mathbb{E}(l_2) = \sum_{i=1}^{n} b_i \mu_i$, then

$$\mathbb{E}(l_1 + l_2) = \sum_{i=1}^{n} (a_i + b_i) \mu_i.$$ 

Now, if $\{x_i\}$ is a random sample with $\mu_{11} = \mu_1$, $\sigma^2 = \sigma^2$, then $\sigma_1^2 = \sigma^2 \sum_{i=1}^{n} a_i^2$, $\sigma_2^2 = \sigma^2 \sum_{i=1}^{n} b_i^2$, and $\sigma_{12} = \sigma^2 \sum_{i=1}^{n} a_i b_i$. Hence the condition that would make $l_1$ and $l_2$ uncorrelated is that $\sum_{i=1}^{n} a_i b_i = 0$. Two uncorrelated linear forms are said to be orthogonal.
Example 6.4. For a random sample of 5 drawn from a normal population with mean, $\mu$, and variance, $\sigma^2$, the mean value and variance of

$$\ell_1 = x_1 + x_2 + x_3 + x_4 + x_5 = \bar{x}$$

are respectively

$$E(\ell_1) = \mu + \mu + \mu + \mu + \mu = 5\mu,$$

$$\sigma^2_\ell = 5\sigma^2.$$

For a second linear form

$$\ell_2 = -2x_1 - x_0 + x_4 + 2x_5$$

we obtain

$$E(\ell_2) = -2\mu - \mu + \mu + 2\mu = 0,$$

also

$$\sigma^2_{\ell_1\ell_2} = (1)(-2) + (1)(-1) + (1)(1) + (1)(2) = 0.$$

For a third linear form

$$\ell_3 = 2x_1 - x_2 - 2x_3 - x_4 + 2x_5,$$

we obtain

$$E(\ell_3) = 2\mu - \mu - 2\mu - \mu + 2\mu = 0,$$

also

$$\sigma^2_{\ell_1\ell_3} = (1)(2) + (1)(-1) + (1)(-2) + (1)(-1) + (1)(2) = 0,$$

and

$$\sigma^2_{\ell_2\ell_3} = (-2)(2) + (-1)(-1) + (0)(-2) + (1)(-1) + (2)(2) = 0.$$

Notice that the sum of the coefficients of $\ell_2$ and $\ell_3$ respectively are zero and that $\ell_2$ and $\ell_3$ are uncorrelated with $\ell_1$, illustrating the theory in the paragraph above.

Example 6.5. To determine a fourth orthogonal linear form,

$$\ell_4 = b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5,$$

to $\ell_1$, $\ell_2$, and $\ell_3$ of Example 6.4 we find
\[ L_1 L_2: \quad b_1 + b_2 + b_3 + b_4 + b_5 = 0 \]
\[ L_2 L_4: \quad -2b_1 - b_2 + b_4 + 2b_5 = 0 \]
\[ L_3 L_4: \quad 2b_1 - b_2 - 2b_3 - b_4 + 2b_5 = 0 \]

Eliminating \( b_3 \) from the first and third equations and then \( b_2 \) from the remaining two equations we find

\[ b_1 + b_4 + 3b_5 = 0 \]

This condition will be satisfied if we let \( b_5 = 1 \), \( b_4 = -2 \), and \( b_1 = -1 \). Then, we find \( b_3 = 2 \) and \( b_2 = 0 \). Hence a fourth orthogonal form desired is

\[ L_4 = -x_1 + 2x_2 - 2x_4 + x_5 \]

There are an infinite number of such functions.

Exercise 6.1. Let the student determine a fifth linear form orthogonal to the four in Examples 6.4 and 6.5. How many linear forms make a complete set?

6.7 - Linear forms with normally distributed variates. Suppose that the \( x_i \) in the linear form \( L = \sum_{i=1}^{n} a_i x_i \) each follow a normal parent population distribution with mean, \( \mu_{x_i} \), and variance, \( \sigma_{x_i}^2 \). Then, the probability of getting the particular \( x_i \)'s in a random sample of \( n \) will be

\[
\frac{1}{(2\pi)^{n/2}} \left( \frac{1}{\sigma_1} \right)^n e^{-\sum_{i=1}^{n} \left( \frac{x_i - \mu_{x_i}}{\sigma_{x_i}} \right)^2} \]

It is desired to find the distribution of \( L - E(L) \).

The m.g.f. of \( L - E(L) \) is

\[ \phi(t) = E \left[ e^{t(L - E(L))} \right] \]
\[-78\]

\[
\frac{1}{(2\pi)^{n/2}} \frac{1}{n^{\alpha_1}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ \frac{1}{\sqrt{2\pi}} \left( \sum_{i=1}^{n} \frac{(x_{1i} - \mu_{1i})^2}{2\sigma_i^2} \right) - t \left( \sum_{i=1}^{n} a_i (x_{1i} - \mu_{1i}) \right) \right\} \prod_{i=1}^{n} dx_i
\]

\[
= \frac{n}{\prod_{i=1}^{n} \sigma_i^{2n}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left( \sum_{i=1}^{n} (x_{1i} - \mu_{1i})^2 / 2\sigma_i^2 + ta_i (x_{1i} - \mu_{1i}) \right) dx_i.
\]

By completing the square of the exponent we have

\[
- \frac{1}{2\sigma_i^2} \left[ (x_{1i} - \mu_{1i})^2 - 2\sigma_i^2 ta_i (x_{1i} - \mu_{1i}) + \sigma_i^2 a_i \right] + \frac{\sigma_i^2 a_i^2}{2}.
\]

\[
= - \frac{1}{2\sigma_i^2} \left[ (x_{1i} - \mu_{1i})^2 - \sigma_i^2 ta_i \right] + \frac{\sigma_i^2 a_i^2}{2}.
\]

Hence,

\[
\phi(t) = \prod_{i=1}^{n} \frac{\sigma_i^2 a_i}{2} = e^{\frac{\sigma_i^2 a_i^2}{2}}.
\]

Now, the mgf for the normal distribution

\[
\frac{1}{\sqrt{2\pi}} e^{-y^2/2\sigma^2} dy, \quad -\infty \leq y \leq \infty,
\]

is

\[
\phi_y(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2\sigma^2 + ty} dy = e^{\frac{\sigma^2 t^2}{2}}.
\]

Hence, we may say that \( L \) is normally distributed with mean \( E(L) = \sum_{i=1}^{n} a_i \mu_{1i} \) and variance \( \sigma_L^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2 \). This result is analogous to that found in Section 6.6 for the mean and variance for non-specified parent population distributions.
References and Other Reading


Exercises

Exercise 6.1 - Given a random sample of \( N \), all members being drawn from the same population. Determine the relationship between the cumulants for the total of the sample and the cumulants for a single member of the sample. Hint: 
\[
\phi_n'(t) = \left[ \phi_1'(t) \right]^N.
\]

Exercise 6.2 - Use the results of Exercise 6.1 to determine the first three cumulants for the total of \( N \) from a binomial distribution. (\( N \) represents the number of samples and \( n \) the number of independent trials per sample.)

Exercise 6.3 - What happens to Exercise 6.1 if each member of the random sample is drawn from different populations? Hint: show that 
\[
K(t) = \sum_{i=1}^{N} K_i(t),
\]

hence, 
\[
K = \sum_{i=1}^{N} K_i.
\]
Consider this problem for Poisson distribution with unlike means and binomial distributions with unlike values of \( n \) and \( p \).

Exercise 6.4 - Given two independent estimates of \( \mu \), \( x_1 \) and \( x_2 \), with variances \( \sigma_1^2 \) and \( \sigma_2^2 \), respectively. If we desire to estimate \( \mu \) as a linear function of the \( x \)'s, find the coefficients of the \( x \)'s which will minimize the variance of the estimate. Remember that \( E(x) = \mu \). Let 
\[
y = \lambda_1 x_1 + \lambda_2 x_2.
\]

Exercise 6.5 - If the \( x_1 \) in the linear form \( \ell = \Sigma a_i x_i \) are normally and independently distributed with means \( \mu \) and variances \( \sigma_1^2 \) and all \( a_i = \frac{1}{N} \), \( \mu_1 = \mu \), and \( \sigma_1 = \sigma \), then \( \ell = \bar{x} \) and we see that the mean of a normal sample is normally distributed with mean \( \mu \) and variance \( \frac{\sigma^2}{N} \).

Exercise 6.6 - Suppose that \( \bar{x}_1 \) represents the mean of a sample of \( n_1 \) taken from a normal population with mean \( \mu_1 \) and variance \( \sigma_1^2 \) and \( \bar{x}_2 \) the mean of a sample of \( n_2 \) taken from another normal population with mean \( \mu_2 \) and variance \( \sigma_2^2 \). Let 
\[
\ell = \bar{x}_1 - \bar{x}_2.
\]
Then we see that \( \ell \) is normally distributed with mean \( \mu_1 - \mu_2 \) and variance
\[
\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.
\]
Exercise 6.7 - Consider the following applied problem. The effect of a nitrogen top-dressing on a crop is to be determined. In addition it is desired to discover at what time during the growing season the top-dressing, if beneficial, should be applied. The experiment was designed as follows: use four plots, one with no nitrogen, one with nitrogen applied early, one applied in the middle of the grow season, and one applied later. Naturally, this entire experiment would be repeated several times, say \( r \), in order to discover how consistent the differences were from replication to replication. The totals of the \( 4r \) plots are indicated as follows:

<table>
<thead>
<tr>
<th>Nitrogen Applied</th>
<th>No N</th>
<th>Early</th>
<th>Middle</th>
<th>Late</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_0 )</td>
<td>( T_1 )</td>
<td>( T_2 )</td>
<td>( T_3 )</td>
<td></td>
</tr>
</tbody>
</table>

where

\[
T_0 = x_{01} + x_{02} + \ldots + x_{0r}; \quad T_1 = \sum_{j=1}^{r} x_{1j}; \quad \text{etc.}
\]

(a) Suppose each \( x \) is \( N(\mu, \sigma^2) \). Set up four pertinent independent linear forms for comparing the four treatments and determine the variance of each.

(b) What happens if we use averages \( \frac{T_1}{r} \) instead of \( T_1 \)?

Exercise 6.8 - Given a series of average yearly prices collected for the years 1935-1939 inclusive. Set up a set of orthogonal linear forms to study the trend in the prices. Can you think of anything which might cause one to alter his method in handling this problem as contrasted to Exercise 6.7 above?
Chapter VII

Derived Sampling Distributions: Normal Parent Population

7.1 - Introduction. In this chapter we shall consider only random samples of \( n \) drawn from a normal parent population with mean \( \mu \) and variance \( \sigma^2 \). If it is desired to include independence, we shall denote this by \( \text{NID}(\mu, \sigma^2) \).

The derived sampling distributions which will be discussed here are of importance in applied statistics in making tests of hypotheses and setting confidence limits. As pointed out earlier, we will assume in these discussions of distribution theory that certain estimates of population parameters and certain tests criteria are the "best" and we will be interested here only in obtaining their probability distributions. In later chapters on tests of hypotheses and estimation a discussion will be given of what properties a "best" estimate or an "appropriate" test criterion should have.

If the specified parent population distribution is not normal, then the derived sampling distributions may become unduly complicated. Non-randomness in the sample presents additional difficulties. Consequences of the relaxation of such assumptions regarding the parent population and the sample have been studied to some extent, and further investigations may be expected in the future.

7.2 - Distribution of the sample mean, \( \bar{x} \). In Section 6.7 it was shown, for a random sample of size \( n \), with each \( x_i \) from a parent population \( N(\mu_i, \sigma_i^2) \), that the distribution of \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \) is \( N(\frac{1}{n} \sum_{i=1}^{n} \mu_i, \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2) \). If \( \sigma_i^2 = \sigma^2 \) and \( \mu_i = \mu \), then \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}, \sum_{i=1}^{n} \mu_i = \mu, \) and \( \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2 = \sigma^2 \).

Therefore the distribution of \( \bar{x} \) is \( N(\mu, \sigma^2/n) \).
Using $\bar{x}$ as an estimate of $\mu$ with the variance of the $x$'s as $\sigma^2$ known, the confidence limits for $\mu$ are

$$
\bar{x} - \frac{t_\alpha \sigma}{\sqrt{n}} < \mu < \bar{x} + \frac{t_\alpha \sigma}{\sqrt{n}},
$$

where $t_\alpha$ is a normal deviate such that

$$
\int_{-t_\alpha}^{t_\alpha} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = 1 - \alpha.
$$

In other words, $t_\alpha$ is so chosen that Prob. $(y > t_\alpha) = \frac{\alpha}{2}$, where $y$ is $N(0,1)$.

Exercise 7.1. If a random sample of size $n_1$ be drawn from $N(\mu_1, \sigma_1^2)$ and a second random sample of size $n_2$ be drawn from $N(\mu_2, \sigma_2^2)$, show that the difference between the two sample means is distributed as

$$
N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}).
$$

Exercise 7.2. (a) Set up the joint probability distribution of a sample of size $n$ from a bivariate normal distribution $(x_1, x_2)$ with means $\mu_1$ and $\mu_2$, variances $\sigma_1^2$ and $\sigma_2^2$ and correlation $\rho$. (b) Show that the simultaneous distribution of $\bar{x}_1$ and $\bar{x}_2$ is

$$
\frac{\sqrt{n}}{2\sigma_1 \sigma_2 \sqrt{1-\rho^2}} e^{-\frac{\theta}{2}} dx_1 dx_2,
$$

where

$$
\theta = \frac{n}{2} \left[ \frac{(\bar{x}_1 - \mu_1)^2}{\sigma_1^2} + \frac{(\bar{x}_2 - \mu_2)^2}{\sigma_2^2} - \frac{2 \rho (\bar{x}_1 - \mu_1)(\bar{x}_2 - \mu_2)}{\sigma_1 \sigma_2} \right],
$$

and

$$
-\infty \leq \bar{x}_1 \leq \infty, \quad -\infty \leq \bar{x}_2 \leq \infty.
$$

7.3 - Chi-square distribution. An important distribution that enters into the theory of derived sampling distributions is the chi-square distribution, defined as
\[ f_\chi^2(\chi^2) d(\chi^2) = \frac{1}{\Gamma(\frac{\nu}{2})} \chi^{\frac{\nu}{2} - 1} e^{-\frac{\chi^2}{2}} d(\chi^2), \quad 0 \leq \chi^2 \leq \infty \]

where \( \nu \) is called the degrees of freedom. It is easy to see that

\[
\int_0^\chi^2 f_\chi^2(\chi^2) d(\chi^2) = \frac{1}{\Gamma(\frac{\nu}{2})} \int_0^{\frac{\chi^2}{2}} y^{\frac{\nu}{2} - 1} e^{-y} dy
\]

which is the Incomplete Gamma Function. This function has been tabulated by Karl Pearson (1) for various values of \( \chi^2 \) and \( \nu \).

Fisher and Yates (2) give values of \( \chi^2_0 \) for \( P = .99, .95, .90, .10, .05, .02, .01, \) and .001, \( \nu = 1 \) (1) 30 where

\[ P = P(\chi^2 > \chi^2_0) = \int_\chi^2_0^\infty f_\chi^2(\chi^2) d(\chi^2). \]

These inverted form tables have been reproduced in most texts on statistical methods. There are many uses of \( \chi^2 \) as a test criterion in making tests of statistical hypotheses. The theory related to such uses will be discussed later.

7.4 - Properties of the chi-square distribution. The m.g.f. for \( \chi^2 \) is

\[ \phi(t) = \frac{1}{\Gamma(\frac{\nu}{2})} \int_0^\infty \left( \frac{\chi^2}{2} \right)^{\nu - 1} e^{-t \frac{\chi^2}{2}} d(\chi^2) \]

\[ = (1 - 2t)^{-\frac{\nu}{2}}. \]

It follows that the cumulant generating function is

\[ K(t) = -\frac{\nu}{2} \log(1 - 2t) = \frac{\nu}{2} \sum_{i=1}^\infty (2t)^i / i, \]
where
\[ k_1 = \sqrt{y} (1-1)! 2^{i-1}. \]

The chi-square distribution has mean \( \sqrt{y} \) and variance \( 2\sqrt{y} \).

Consider a new variable \( y = \frac{\chi^2 - \sqrt{y}}{2\sqrt{y}} \), for which \( \mu = 0 \), \( \sigma^2 = 1 \).

The cumulant generating function for this new variable is
\[ K(t) = \frac{\sqrt{y}}{2} \sum_{i=2}^{\infty} \frac{(2t)^i}{i(2\sqrt{y})^{1/2}}, \]

and
\[ k_1 = (2/\sqrt{y}) \frac{i-1}{2} (1-1)! , \quad i \geq 2. \]

Then
\[ k_3 = \frac{2\sqrt{2}}{\sqrt{y}} , k_4 = \frac{12}{y} , \text{ etc.} \]

Hence, for \( i \geq 2 \)
\[ \lim_{\sqrt{y} \to \infty} k_i = 0. \]

But this is a property of the normal distribution indicating that the chi-square distribution approaches the normal distribution for large \( \sqrt{y} \).

**Exercise 7.3.** Given the distribution of \( \chi^2 \) as in the Section 7.3, find the distribution of \( y = \frac{\chi^2}{1+\chi^2} \).

**Exercise 7.4.** Given the two functions
\[ u = a_1 \chi_1^2 + a_2 \chi_2^2 \]
\[ v = \chi_1^2 + \chi_2^2 \]

Find the distribution of \( W = \frac{u}{v} \). What effect would the provision \( a_1 > a_2 > 0 \) have on the results?

**Exercise 7.5.** If \( x_1 \) and \( x_2 \) are NID(0,1), show that the variables \( x_1 \) and \( \theta \) defined as follows:
\[ x_1 = r \sin \theta, \quad x_2 = r \cos \theta \]
are independently distributed and that \( r^2 \) is distributed as \( \chi^2 \) with 2 degrees of freedom.
7.5 - Distribution of a sum of squares of deviations. Given a sample \( \{ x_i \} \) of \( n \) from the population of \( x \)'s which are \( \operatorname{NID}(\mu, \sigma^2) \). Then the probability of obtaining this particular sample of \( x \)'s or the joint probability distribution is

\[
f \left( x_1^2, \ldots, x_n^2 \right) = \frac{1}{\sigma^{2n} 2^n \pi^{n/2}} e^{-\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2 \sigma^2}} dx_1 \ldots dx_n.
\]

Assume that \( \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{\sigma^2} = \chi^2 \) with \( \nu = n \) degrees of freedom. The m.g.f. for this \( \chi^2 \) is

\[
\phi'(t) = \left( \frac{1}{\sigma \sqrt{2\pi}} \right)^n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{tx_1^2 - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2 \sigma^2}} dx_1 \ldots dx_n
\]

\[
= (1 - 2t)^{-n/2}
\]

But this is the m.g.f. of a \( \chi^2 \) with \( \nu = n \) degrees of freedom, hence the assumption that \( \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{\sigma^2} \) is distributed as \( \chi^2 \) with \( \nu = n \) degrees of freedom is correct. In this case each \( x \) furnishes a single degree of freedom, hence the number of degrees of freedom is the number of independent values which make up \( \chi^2 \). The number of degrees of freedom in general is the total number of observations less the number of independent restraints imposed on the observations in forming the distribution.

7.6 - Reproductive property of the chi-square distribution. Using the relationship \( \chi^2 = \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2} \) we may prove that the sum of \( n \) variates, each independently distributed as \( \chi^2 \) with \( \nu_i \) degrees of freedom, is itself distributed as \( \chi^2 \) with \( \nu = \sum_{i=1}^{n} \nu_i \) degrees of freedom. Suppose
\[ \chi^2_i = \frac{\gamma_1}{\Sigma} \left( \frac{\Sigma}{\Sigma} \right) \left( x_{1i} - \mu_1 \right)^2 \]

and

\[ \chi^2_3 = \frac{\gamma_2}{\Sigma} \left( \frac{\Sigma}{\Sigma} \right) \left( x_{2i} - \mu_2 \right)^2. \]

Then

\[ \chi^2 = \chi^2_1 + \chi^2_3 \]

has the m.g.f.

\[ \phi'(t) = \left( \frac{1}{c_1 \sqrt{2\pi}} \right) ^{\gamma_1} \left( \frac{1}{c_2 \sqrt{2\pi}} \right) ^{\gamma_2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\phi} \frac{\gamma_1}{\gamma_1} \frac{\gamma_2}{\gamma_2} dx_{1i} dx_{2i} \]

where

\[ \phi = \left[ \frac{\gamma_1}{\Sigma} \left( x_{1i} - \mu_1 \right)^2 + \frac{\gamma_2}{\Sigma} \left( x_{2i} - \mu_2 \right)^2 \right] \left( 1 - 2t \right). \]

Performing the integration we find

\[ \phi'(t) = (1 - 2t) \frac{\gamma_1 + \gamma_2}{2}. \]

But this is the m.g.f. for a \( \chi^2 \) with \( \gamma_1 + \gamma_2 \) degrees of freedom, hence \( \chi^2_1 + \chi^2_3 \) is distributed similarly. This result may be extended to \( n \) variates.

Exercise 7.5. Given the bivariate normal distribution with means zero and variances \( c_1^2 \) and \( c_2^2 \) and correlation \( \rho \). Show that the quadratic form

\[ \theta = \frac{x_1^2}{c_1^2} + \frac{x_2^2}{c_2^2} - \frac{2 \rho x_1 x_2}{c_1 c_2} \]

is distributed as \( \chi^2 \) with 2 degrees of freedom. Hint: Show that the m.g.f. of \( \theta \) is the same as that of \( \chi^2 \).
Exercise 7.6. Using the result in Exercise 7.5 find the probability that \( \Theta \leq \sigma^2 \).

Exercise 7.7. In Exercise 7.5 let \( \sigma_1^2 = \sigma_2^2 = 1 \) and \( \rho = 0 \). Find the equation of the circle with center at the origin which would contain 99 percent of a large number of samples of \( x_1 \) and \( x_2 \).

Exercise 7.8. A random sample of \( n \) individuals \( \{x_i\} \) is drawn from a \( N(0,1) \) population. Suppose the sample is subdivided into \( r \) subclasses with 2 or more in each \( (n_1, n_2, \ldots, n_r \) being the subclass members). The sum of squares \( q_i (i = 1, 2, \ldots, r) \) of deviations from the sub-class mean is computed for each sub-class. What is the distribution of \( \sum_{i=1}^{r} q_i^2 \)?

Exercise 7.9. Compute the 1% and 5% significance levels of \( \chi^2 \) for \( \gamma = 1, 2, \) and \( 4 \). Check with the results in the statistical tables.

Exercise 7.10. Given \( \frac{(s^2)^g}{\sigma^2} = 2 \) with \( n = 4 \), where \( (s^2)^g = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} \). What is the probability of obtaining a value this large or larger. Check your results by interpolating with the tabled values.

7.7 - Simultaneous distribution of the sample mean \( \bar{x} \) and variance estimates \( \sigma^2 \). It was shown in Section 7.5 that \( \sum_{i=1}^{n} (x_i - \mu)^2 \) is distributed as \( \chi^2 \) with \( \gamma = n \) degrees of freedom. We shall now obtain the simultaneous distribution of two parts of this sum when expressed as

\[
\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sigma^2} + \frac{n(\bar{x} - \mu)^2}{\sigma^2}
\]

\[
= \frac{(n-1)s^2}{\sigma^2} + \frac{n(\bar{x} - \mu)^2}{\sigma^2},
\]

where it will be recalled that \( \bar{x} \) is an unbiased estimate of \( \mu \) and \( s^2 \) is an unbiased estimate of \( \sigma^2 \).
In order to simplify the notation in the following derivation, let 
\[ u = \frac{X - \mu}{\sigma} \], so that \( u \) is \( N(0,1) \) and \( \tilde{u} = \frac{X - \mu}{\sigma} \). Let \( \gamma^2 = \frac{\sigma^2}{\sigma^2} \).

The derivation will be accomplished by making use of orthogonal linear
forms as discussed in Section 6.6. To this end we set up the following
orthogonal linear transformations:

\[
y_1 = \frac{u_1 + u_2 + \ldots + u_n}{\sqrt{n}}
\]

\[
y_2 = \frac{u_1 - u_2}{\sqrt{2}}
\]

\[
y_3 = \frac{u_1 + u_2 - 2u_3}{\sqrt{6}}
\]

\[
\vdots
\]

\[
y_n = \frac{u_1 + u_2 + \ldots + u_{n-1} - (n-1)u_n}{\sqrt{n(n-1)}}
\]

It is easily seen that each \( y_1 \) is orthogonal to each \( y_j \) \((i \neq j)\). Also
the denominators have been chosen to make the variances of the \( y \)'s equal
unity. This type of transformation is called completely orthogonal and
these requirements may be designated as

\[
E(y_iy_j) = \begin{cases} 
0 & i \neq j \\
1 & i = j 
\end{cases}
\]

From Section 6.7 we know that the \( y_i \)'s are \( NID(0,1) \); hence, the simultaneouse distribution of the \( \{y_i\} \) is identical with that of the \( \{u_i\} \) and

\[
\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} u_i^2. \quad \text{Also since } y_1^2 = n\tilde{u}^2, \text{ then}
\]

\[
\sum_{i=1}^{n} u_i^2 = (n-1)\gamma^2.
\]

Since \( y_1 \) is independent of the other \( y \)'s, \( \tilde{u} \) is independent of \( \gamma^2 \). Hence

\[
f(\tilde{u},\gamma^2)d\tilde{u}(\gamma^2) = f_1(\tilde{u})f_2(\gamma^2)d\tilde{u}d(\gamma^2)
\]

We know that

\[
f_1(\tilde{u})d\tilde{u} = \sqrt{\frac{2}{\pi\tilde{u}^2}} e^{-\frac{\tilde{u}^2}{2}} d\tilde{u}.
\]
It is also known that
\[ \sum_{i=2}^{n} y_i^2 \]
is distributed as \( \chi^2 \) with \( (n-1) \) degrees of freedom since there are \( (n-1) \) independent variates. Now, \( \chi^2 = (n-1)\gamma^2 \), hence
\[
f_\chi(\gamma^2) d(\gamma^2) = \left( \frac{n-1}{2} \right)^{\frac{n-1}{2}} \left( \gamma^2 \right)^{\frac{n-3}{2}} e^{-\frac{(n-1)\gamma^2}{2}} \Gamma \left( \frac{n-1}{2} \right) d(\gamma^2).
\]
Using the original \( x \)'s, which were \( N(\mu, \sigma^2) \), the simultaneous distribution of \( \bar{x} \) and \( s^2 \) is
\[
f(\bar{x}, s^2) d\bar{x} d(s^2) = \frac{\sqrt{3}}{\sqrt{2\pi}} \left( \frac{n-1}{2} \right)^{\frac{n-1}{2}} \frac{n-3}{2} e^{-\frac{1}{2} \left( \frac{\left( \bar{x} - \mu \right)^2 + (n-1)s^2}{\sigma^2} \right)} d\bar{x} d(s^2)
\]

Exercise 7.11. Show that the distribution of \( s^2 \) approaches normality for large \( n \). What are the skewness and kurtosis of \( s^2 \)?

Exercise 7.12. Find the m.g.f. of \( \log s \), where \( s^2 \) is an estimate of variance. Show that the distribution of \( \log s \) is independent of \( \sigma^2 \) apart from its mean value.

Exercise 7.13. Given \( \frac{s^2}{\sigma^2} = 2 \) with \( n = 4 \), where \( s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 \). What is the probability of obtaining a value this large or larger? Check your results by interpolating with the tabulated values.

7.8 - Distribution of \( t \). It is now proposed to obtain the distribution of a test criterion to be used in testing the hypothesis that a sample \( \{x_i\} \) was drawn from a normal population with a specified mean \( \mu \). The distribution must be independent of \( \sigma^2 \), if it is to be of practical use, since we seldom know the value of \( \sigma^2 \).
One proposed test criterion of this hypothesis is

\[ t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \]

where \( \bar{x} \) and \( s \) are computed from a random sample of \( n \) from the parent population \( N(\mu, \sigma^2) \). Two properties of \( t \) should be noted: (i) \( t \) is the ratio between a normal deviate and the square root of an unbiased estimate of its variance. (ii) \( t^2 \) is the ratio between the square of a variate, \( y = \frac{n(\bar{x} - \mu)}{\sigma} \), which is \( N(0, 1) \) and a variate \( \frac{s^2}{\sigma^2} \), which is distributed as \( \chi^2_n \) with \( n \) degrees of freedom. Note that \( \sigma^2 \) cancels in the derivation of \( t \).

From Section 7.7

\[ f(\bar{x}, s^2)d\bar{x}ds^2 = k\left(\frac{s^2}{\sigma^2}\right)^{n-1} e^{-\left[\frac{n(\bar{x} - \mu)^2 + (n-1)s^2}{2\sigma^2}\right]} d\bar{x}d(s^2) \]

where

\[ k = \frac{\sqrt{\frac{\pi}{2n}}} {\left[\frac{\Gamma\left(\frac{2n}{2}\right)}{2\Gamma\left(\frac{n}{2}\right)}\right]} \]

Writing \( f(\bar{x}, s^2) \) as a function of \( t \) and \( s^2 \) we obtain,

\[ f(t, s^2) = k\left(\frac{s^2}{\sigma^2}\right)^{\frac{n-2}{2}} e^{-\frac{(n-1)s^2(1 + \frac{t^2}{n-1})}{2\sigma^2}} ds^2 dt \]

Let

\[ u = \frac{(n-1)s^2}{2\sigma^2} \left(1 + \frac{t^2}{n-1}\right) \]

then

\[ f_{n-1}(t) = k u^{\frac{1}{2}} \Gamma\left(\frac{n}{2}\right) e^{-\frac{u}{2}} du \int_0^\infty \frac{u^{\frac{n-1}{2}}}{2}\ e^{-u} du \]

\[ = k u^{\frac{1}{2}} \Gamma\left(\frac{n}{2}\right) \left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}} dt, \quad -\infty \leq t \leq \infty. \]
Since the area under the curve must be 1 for a probability distribution and because of symmetry we have

\[ k^{11} \int_{0}^{\infty} \left( 1 + \frac{t^2}{n-1} \right)^{-\frac{n}{2}} dt = \frac{1}{2} \]

Let \( t^2 = (n-1)u \),

then

\[ k^{11} \int_{0}^{\infty} (n-1)^{1/2} (1+u)^{-\frac{1}{2}} du = 1 \]

Let \( y = \frac{u}{1+u} \),

then

\[ k^{11} \int_{0}^{1} y^{1/2} \left( 1-y \right)^{\frac{n-1}{2}} dy = 1. \]

Hence

\[ k^{11} \int_{0}^{\infty} (n-1)^{1/2} B\left(\frac{1}{2}, \frac{n-1}{2}\right) = 1. \]

and

\[ k^{11} = \frac{1}{\sqrt{\pi(n-1) B\left(\frac{1}{2}, \frac{n-1}{2}\right)}} \]

The distribution becomes

\[ f_{n-1}(t)dt = \frac{\int_{0}^{\infty} \left( 1 + \frac{t^2}{n-1} \right)^{-\frac{n}{2}} dt}{\sqrt{\pi(n-1) B\left(\frac{1}{2}, \frac{n-1}{2}\right)}} \]

\[ -\infty \leq t \leq \infty. \]

Since \( s^2 \) has \( (n-1) \) degrees of freedom, \( f_{n-1}(t)dt \) is designated as the distribution of \( t \) for \( (n-1) \) degrees of freedom. The distribution for \( n \) degrees of freedom is

\[ f_n(t)dt = \frac{\int_{0}^{\infty} \left( 1 + \frac{t^2}{n} \right)^{-\frac{n+1}{2}} dt}{\sqrt{\pi n B\left(\frac{1}{2}, \frac{n+1}{2}\right)}} \]

\[ -\infty \leq t \leq \infty. \]

It should be noted that

\[ P\left[ |t| \geq t_0 \right] = 2 \int_{t_0}^{\infty} f_n(t)dt. \]
Tables of values of $t_0$ have been prepared for various values of $n$ and $P$ with more complete values for $P = .05$ and .01. These $t$-tables can be found in most modern texts on statistical methods. Note that the probabilities given in the tables are for a two-tailed test. If it is desired to obtain the probability that $t > t_0$, the tabled probabilities must be divided by 2.

If it is assumed that $\{x_i\}$ is a random sample from a normal population with a specified mean $\mu$, then $P$ is the probability of obtaining a value of $t$ which diverges in absolute value by as much as or more than $t_0$ from the mean above stated value, $t=0$. As used in testing the hypothesis, $P$ is called the significance level for testing the hypothesis that the sample was drawn from a normal population with specified mean $\mu$. If $P$ is small, we conclude that it is unlikely that the specified population mean $\mu$ is the actual population mean.

If we use $\bar{x}$ as an estimate of $\mu$ and $s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$ as an estimate of $s^2$, the two-tailed confidence limits for $\mu$ are

$$\bar{x} - t_{\alpha} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha} \frac{s}{\sqrt{n}},$$

when the Prob. ($t > t_{\alpha}$) = $\frac{\alpha}{2}$.

7.9 - Test of the difference between two means. Another important use of the $t$-distribution is in testing an hypothesis concerning the difference of two population means. Suppose that we have two random samples $x_{1i}$ and $x_{2i}$ from populations $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively with $\sigma_1^2 = \sigma_2^2 = \sigma^2$. We wish to test the hypothesis that $\mu_1 = \mu_2$. The procedure in constructing a suitable test criterion will be to: (i) find a statistic, independent of $\sigma^2$, which involves the differences between the population means $\mu_1$ and $\mu_2$, (ii) find the sampling distribution of the statistics on the assumption that the hypothesis tested is true, i.e. in this case, $\mu_1 = \mu_2$. 
Let the samples have the following characteristics:

<table>
<thead>
<tr>
<th>Sample</th>
<th>Size</th>
<th>Mean</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>{x_{11}}</td>
<td>n_1</td>
<td>\bar{x}_1</td>
<td>s_1^2</td>
</tr>
<tr>
<td>{x_{21}}</td>
<td>n_2</td>
<td>\bar{x}_2</td>
<td>s_2^2</td>
</tr>
</tbody>
</table>

Now, \(\frac{(n_1-1)s_1^2}{c_1^2}\) is distributed as \(\chi^2\) with \((n_1-1)\) degrees of freedom. Similarly, \(\frac{(n_2-1)s_2^2}{c_2^2}\) is distributed as \(\chi^2\) with \((n_2-1)\) degrees of freedom.

Since \(c_1^2 = c_2^2 = \sigma^2\), \(\frac{(n_1-1)s_1^2}{c_1^2} + \frac{(n_2-1)s_2^2}{c_2^2} = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{\sigma^2}\).

By the reproductive property of \(\chi^2\), assuming the two samples are independently drawn, this quantity is distributed as \(\chi^2\) with \((n_1+n_2-2)\) degrees of freedom. Then,

\[s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}\]

is an unbiased estimate of \(\sigma^2\) since its expected value is \(\sigma^2\). Also the variance of \((\bar{x}_1 - \mu_1) - (\bar{x}_2 - \mu_2)\) is

\[\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} = \sigma^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right).\]

The quantity

\[\frac{(\bar{x}_1 - \mu_1) - (\bar{x}_2 - \mu_2)}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}\]

is the ratio of a normal deviate to the square root of an unbiased estimate of its variance. Writing the quantity as

\[
\frac{\left[(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)\right]^2 / \sigma^2 (\frac{1}{n_1} + \frac{1}{n_2})}{s^2 / \sigma^2}
\]

we see that it is now the ratio of the square of a variate which is \(N(0,1)\) and a variate which is distributed as \(\chi^2\) with \((n_1+n_2-2)\) degrees of freedom. These two properties are those stated in Section 7.5 for \(t\) and \(t^2\) respectively. But it can be proved that, if a quantity possesses these
properties, it is distributed as $t$. Hence, the test criterion on the assumption that the hypothesis tested ($\mu_1 = \mu_2$) is true is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$  

The two-tailed confidence limits for the true difference between the two means, $\mu_1 - \mu_2 = \delta$, are given by

$$d - t_{2\alpha} s_d \leq \delta \leq d + t_{2\alpha} s_d,$$

where $d = \bar{x}_1 - \bar{x}_2$ and $s_d = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$.

If $s_1^2 \neq s_2^2$, then the variance of $(\bar{x}_1 - \mu_1) - (\bar{x}_2 - \mu_2)$ is

$$\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2},$$

for which

$$\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$$

is an unbiased estimate. Hence the first property of $t$ is met by

$$t^2 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

but not the second, since the square of the denominator is not distributed as $\chi^2$, the quantity $m$ being the number of degrees of freedom.

We know that $\frac{(n_1-1)s_1^2}{c_1^2}$ is distributed as $\chi^2$ with $(n_1-1)$ degrees of freedom and similarly for $\frac{(n_2-1)s_2^2}{c_2^2}$. Hence a denominator analogous to that of $t^2$ is

$$\frac{(n_1-1)s_1^2/c_1^2 + (n_2-1)s_2^2/c_2^2}{(n_1 + n_2 - 2)},$$

and an analogous numerator is

$$\frac{[\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)]^2}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$
If we set
\[ t = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \left/ \sqrt{\frac{(n_1-1)\sigma_1^2/\sigma_1^2 + (n_2-1)\sigma_2^2/\sigma_2^2}{n_1 + n_2 - 2}} \right. \]
and let \( \lambda = \frac{\sigma_2^2}{\sigma_1^2} \) and \( \ell = \frac{c_2^2}{\bar{x}_1^2} \), then
\[ t' = \frac{t \sqrt{(n_1-1) + (n_2-1)\ell/\lambda}}{\sqrt{n_1 + n_2 - 2}} \left/ \sqrt{\frac{1}{n_1} + \frac{\lambda}{n_2}} \right. \left/ \sqrt{\frac{1}{n_1} + \frac{\ell}{n_2}} \right. \]
is an analogous expression to the ordinary \( t \). The exact distribution of \( t' \) is not known and even if known would be of little practical use since it would undoubtedly involve the population parameters \( \sigma_1^2 \) and \( \sigma_2^2 \).

Two methods have been devised for using \( t' \) to make the desired test. They are described by Bartlett (3), (4), and Welch (5).

Exercise 7.14. The distribution of \( t \) for 1 degrees of freedom is the so-called Cauchy distribution. What are its mean and variance? Compute the 1% and 5% significance levels of this \( t \), using the two-tailed test.

Exercise 7.15. (a) Given a random sample of \( n \) paired observations (A and B). The sample mean of the A's is \( \bar{X}_a \) and that of the B's is \( \bar{X}_b \). The A's and B's come from normal populations with the same variance. It is desired to test whether the population means are equal. The value of the \( i^{th} \) members of A and B may be represented respectively as
\[ A_i = \mu_a + X_i + e_{ai}, \]
\[ B_i = \mu_b + X_i + e_{bi}, \]
where \( \mu_a \) and \( \mu_b \) are the population mean effects of A and B, \( X_i \) represents the effect common to the two paired observations, assumed \( N(0, \sigma_p^2) \). The \( e_{ai} \) and \( e_{bi} \) represent residual effects for each observation not explained by the \( \mu \)'s and \( X \) and independent of them. The \( e_{ai} \) and \( e_{bi} \) are assumed \( N(0, \sigma^2) \).
If
\[ \bar{x}_a = \frac{1}{n} \sum_{i=1}^{n} A_i , \quad \bar{x}_b = \frac{1}{n} \sum_{i=1}^{n} B_i \]
and
\[ s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left( (A_i - B_i) - (\bar{x}_a - \bar{x}_b) \right)^2 \]
show that
\[ \frac{(\bar{x}_a - \bar{x}_b) \sqrt{n}}{s} \]
follows the \( t \) distribution.

b. Compare \( s^2 \) in (a) with that obtained by pooling the two sample sums of squares of deviations.

Exercise 7.16. Show that the \( t \)-distribution approaches normality for large \( n \).

Exercise 7.17. The distribution of the correlation coefficient \( r \) in samples of \( n \) from a normal bivariate population with \( \rho = 0 \) is

\[ f(r) = \frac{1}{\pi B\left(\frac{n-2}{2}, \frac{1}{2}\right)} \left(1-r^2\right)^{\frac{n-4}{2}} \text{ for } 0 \leq r \leq 1. \]

Show that
\[ r \sqrt{(n-2)/(1-r^2)} \]
is distributed as \( t \) with \( (n-2) \) degrees of freedom.

Exercise 7.18. Compute the 5\% significance level of \( t \) (two-tailed test) for \( n = 3 \). What is the probability that \( t > 3 \)?

Exercise 7.19. In a certain test, given to the 45 members of a class in statistics, 20 women students had an average score of 40 with a variance of 16, while 25 male students had an average of 46 with a variance of 16.
What is the probability of obtaining such results if both groups were equally well prepared for the test? What assumptions are made in obtaining this probability value? On the basis of this sample and the validity of the assumptions is there much evidence that both groups are not equally well prepared?

Exercise 7.20. Given the following data on the gains (in lbs.) by 27 hogs, in individual and similar pens, 12 fed ration A and 15 ration B. Test the hypothesis that the population mean gains are the same.

A 25, 30, 28, 34, 24, 25, 13, 32, 24, 30, 31, 35
B 44, 34, 22, 8, 47, 31, 40, 30, 32, 35, 18, 21, 35, 29, 22

Exercise 7.21. In a certain school 3½ year beginners in the first grade, were selected and paired on the basis of I.Q., socio-economic rating of family, general health, and family size. One member of each pair had attended one year of kindergarten while the other had not. On a certain first grade readiness test given to all 3½ pupils the scores were

Kindergarten 65, 66, 70, 63, 61, 62, 73, 75, 72, 78, 64, 73, 79, 80, 67, 74, 82
No Kindergarten 65, 68, 65, 60, 65, 60, 72, 75, 73, 70, 66, 70, 77, 76, 63, 74, 78

Is there significant evidence from this investigation that kindergarten is of benefit in preparing for the first grade? In making use of the t-test what null hypothesis is made concerning \( \mu_1 = \mu_2 \)? What must be true concerning the distribution of the paired differences?

7.10 Distribution of \( \chi^2 \). Another very useful statistic in making tests of hypotheses involves the ratio of two chi-squares. Let \( \chi^2_1 \) and \( \chi^2_2 \) be independently distributed with \( n_1 \) and \( n_2 \) degrees of freedom, respectively. Then their simultaneous or joint distribution is
\[ f(\chi_1^2, \chi_2^2) \, d(\chi_1^2) \, d(\chi_2^2) = \frac{1}{\Gamma(\frac{n_1}{2}) \Gamma(\frac{n_2}{2})} \left( \frac{\chi_1^2 + \chi_2^2}{2} \right)^{\frac{n_1+n_2-2}{2}} e^{-\frac{\chi_1^2}{2} - \frac{\chi_2^2}{2}} \, d(\chi_1^2) \, d(\chi_2^2) \]

Let

\[ F = \frac{n_1 \chi_1^2}{n_1 + n_2} \quad \text{or} \quad \chi_1^2 = \frac{n_1}{n_2} F, \]

Then

\[ f(F, \chi_2^2) = \frac{\frac{n_1}{2}}{2 \Gamma(\frac{n_1}{2}) \Gamma(\frac{n_2}{2})} \left( \frac{\chi_2^2}{2} \right)^{\frac{n_1+n_2-2}{2}} e^{-\frac{\chi_2^2}{2}} \frac{n_1+n_2}{n_2} \left( \frac{1}{n_2} F \right)^{-\frac{n_1+n_2}{2}}, \]

and

\[ f(F) \, dF = \left[ \int_0^\infty f(F, \chi_2^2) \, d(\chi_2^2) \right] \, dF \]

\[ = C F^{\frac{n_1}{2} - 1} \left( 1 + \frac{n_1}{n_2} F \right)^{-\frac{n_1+n_2}{2}} \, dF, \quad 0 \leq F \leq \infty \]

Exercise 7.22. Show that

\[ C = \frac{\left( \frac{n_1}{n_2} \right)^{\frac{n_1}{2}}}{B \left( \frac{n_1}{2}, \frac{n_2}{2} \right)} \]

Exercise 7.23. Show that

\[ \int_0^\infty f(F) \, dF = 1 \]

Exercise 7.24. When \( n_1 = 1 \), show that \( F = t^2 \).

Since \( n_1 \) and \( n_2 \) are degrees of freedom

\[ \chi_1^2 = \frac{n_1}{c_1^2} \quad \text{and} \quad \chi_2^2 = \frac{n_2}{c_2^2} \]

where \( c_1^2 \) and \( c_2^2 \) are two independent sample estimates of \( c_1^2 \) and \( c_2^2 \), respectively. Then

\[ F = \frac{c_1^2}{c_2^2} \cdot \frac{\chi_2^2}{\chi_1^2} \cdot \frac{\chi_1^2}{\chi_2^2} \]
This statistic provides a test of the hypothesis \( s_1^2 = s_2^2 = s^2 \). Then

\[
F = \frac{s_1^2}{s_2^2} = F_0,
\]

which is the ratio of two sample variances, each assumed to be an independent estimate of the same population variance.

The probability integral for the \( F \)-distribution

\[
P(F \geq F_0) = \int_{F_0}^{\infty} f(F) dF
\]

gives the probability that a value of \( F \) as large as or larger than the computed \( F_0 \) could have been obtained from two random samples from the same population. If \( F \) is small we may choose to reject the hypothesis, knowing that the probability of being wrong, in such case, is equal to or less than \( F \). In practice, we decide that we are willing to take a certain risk in rejecting the hypothesis, say \( F = .05 \), when the hypothesis is actually true. \( F \) is called the probability of making an error of the first kind – Type I Error. By calculation or from tables we obtain the value of \( F_0 \) corresponding to \( F = .05 \) and the particular values of \( n_1 \) and \( n_2 \). If the value obtained from the sample of \( F = \frac{s_1^2}{s_2^2} \) is equal to or greater than \( F .05 \), we reject the hypothesis at the 5% probability level.

If we fail to reject the hypothesis, on the basis of the above procedure, when in fact it is false, we are making an error of the second kind – Type II Error. In obtaining a statistic to test an hypothesis it is desirable to select the significance region so as to minimize the probability of making a Type II error. The theory behind these concepts has been developed by Neyman and Pearson and extended by Wald, Woldowitz, Wilks and others. It will be discussed in later chapters.
In the above discussions we have been assuming that the only alternative to \( \sigma_1^2 = \sigma_2^2 \) is that \( \sigma_1^2 > \sigma_2^2 \). Of course, even if \( \sigma_1^2 > \sigma_2^2 \), the sample \( F_0 \) might be less than 1, but this would only increase our confidence in the truth of the hypothesis that \( \sigma_1^2 = \sigma_2^2 \). In fact it can be shown that if we use the region \( F > F_0 \) to test the hypothesis \( \sigma_1^2 = \sigma_2^2 \) against the single alternative \( \sigma_1^2 > \sigma_2^2 \), we will minimize the chance of making a Type II Error.

Under certain conditions of an applied problem it may not be valid to assume that the only alternative to the hypothesis \( \sigma_1^2 = \sigma_2^2 \) is \( \sigma_1^2 > \sigma_2^2 \). If \( \sigma_1^2 \) may also be less than \( \sigma_2^2 \), then the alternative hypothesis is \( \sigma_1^2 \neq \sigma_2^2 \), and it becomes necessary to consider both tails of the \( F \)-distribution in testing the hypothesis. In this case

\[
P = 2 \int_{F_0}^{\infty} f(F)\,dF
\]

where \( F \) is defined as the larger mean square divided by the smaller.

In Section 7.9 it was shown that \( t \) may be used to test the hypothesis that \( \mu_1 = \mu_2 \) on the assumption that \( \sigma_1^2 = \sigma_2^2 \). Suppose that for the two samples \( \{x_{11}\} \) and \( \{x_{21}\} \) there is some reason to believe that the population variances are not equal but there is no a priori reason for one to be larger than the other. In this case it is necessary to use the two-tailed \( F \)-test in testing for the significance of the difference between the two variances. On the other hand in applying the \( F \)-test to an analysis of variance table, to be explained later, the usual alternative hypothesis is \( \sigma_1^2 > \sigma_2^2 \) and a single-tailed \( F \)-test is used.

Tables of \( F_{.05} \) and \( F_{.01} \) have been compiled by Snedecor in Statistical Methods and have been reproduced in most modern texts on statistical methods. Fison has computed values of \( F_{.20} \) and Colcord and Deming of \( F_{.001} \). All of these tables may be found in Fisher and Yates, Statistical Tables. These significance levels are for single-tailed probabilities \( (\sigma_1^2 > \sigma_2^2) \). If a
two-tailed test is required, these probabilities must be doubled. For example, \( F_{.05} \) now becomes \( F_{.10} \). Tables of the \( F \)-distribution have also been compiled by Merrington and Thompson in Biometrika 32(1941) for \( P = .005, .01, .025, .05, .10, .25, \) and .50.

Exercise 7.25. Derive the distribution of \( F' = \frac{1}{F} \). Hence show that

\[
P \left[ F(n_2, n_1) < F_0 \right] = P \left[ F(n_1, n_2) > \frac{1}{F_0} \right],
\]

where \( F(n_2, n_1) \) indicates there are \( n_2 \) degrees of freedom for the numerator and \( n_1 \) for the denominator. Hence show that for the two-tailed test,

\[
P = 2 \int_{F_0}^{\infty} f(F) \, dF
\]

where \( F \) is the larger mean square divided by the smaller.

Exercise 7.26. Show that \( P(F > F_0) \) is an Incomplete Beta Function. Set up this function.

Exercise 7.27. Use the result in Exercise 7.26 to determine the general formula for \( \mu_1' \) of \( F \).

Exercise 7.28. Determine the 5% significance level (single-tail) for \( F \) if (a) \( n_1 = 2, n_2 = 4 \), and (b) \( n_1 = 4, n_2 = 4 \).

Exercise 7.29. What is \( P(F > 4) \) for \( n_1 = 4, n_2 = 4 \)?

Exercise 7.30. Given \( s_1^2 = 40, s_2^2 = 10 \) with \( n_1 = 4, n_2 = 4 \). Determine the probability that sample variances as divergent as these could be estimates of the sample population variance with alternative hypothesis that \( c_2^2 \neq c_3^2 \).
Exercise 7.31. Given that one mean square, 30, with 4 degrees of freedom is an estimate of $\sigma_3^2 + 5\sigma_1^2$ while another mean square, 10, with 20 degrees of freedom is an estimate of $\sigma_2^2$. Use an $F$-table to test the hypothesis that $\sigma_1^2 = 0$.

Exercise 7.32. R. A. Fisher first considered the problem of the ratio of two variances as the difference between the logarithms of the two variances. If we let

$$z = \frac{1}{2} \log_e F,$$

we obtain his $z$-distribution, which is more nearly normal than the $F$-distribution. Show that

$$f(z)dz = 2e^{\frac{2z}{2}} \left(1 + \frac{n_1}{n_2} \cdot e^{-2z} \right) dz,$$

$$0 \leq z \leq \infty.$$

Exercise 7.33. Show that $f(z)$ is symmetrical if $n_1 = n_2$. 
References and Other Reading


(5) Welch, B. L. Significance of the difference between two means when the population variances are unequal. Biometrika 29. (1935).

PART II

THE THEORY AND APPLICATION OF ESTIMATION BY LEAST SQUARES
<table>
<thead>
<tr>
<th>Page</th>
<th>line*</th>
<th>At present</th>
<th>Should be</th>
</tr>
</thead>
<tbody>
<tr>
<td>105</td>
<td>9</td>
<td>$n(\mu, \sigma^2)$</td>
<td>$N(\mu, \sigma^2)$</td>
</tr>
<tr>
<td>108</td>
<td>1</td>
<td>$Y$</td>
<td>$\hat{Y}$</td>
</tr>
<tr>
<td>109</td>
<td>13</td>
<td>$Y = SY/n = \mu + \varepsilon_0$</td>
<td>$ar{Y} = SY/n = \mu + \varepsilon_0$</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>$E(\varepsilon_0)^2$</td>
<td>$E(\varepsilon_0)^2$</td>
</tr>
<tr>
<td>113</td>
<td>1</td>
<td>$\sqrt[11]{\frac{Sx_1^2}{s_1^2}}$</td>
<td>$\sqrt[11]{\frac{Sx_1^2}{s_1^2}}$</td>
</tr>
<tr>
<td>119</td>
<td>9</td>
<td>$Y$ or $X_1$</td>
<td>$Y$ on $X_1$</td>
</tr>
<tr>
<td>120</td>
<td>9b</td>
<td>DeLury</td>
<td>DeLury (9)</td>
</tr>
<tr>
<td>122</td>
<td>6b</td>
<td>$S, s_1, s_2$</td>
<td>$G, s_1, s_2$</td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>$S$</td>
<td>$G$</td>
</tr>
<tr>
<td>124</td>
<td>9</td>
<td></td>
<td>Add $1/$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$1/x_2 - \frac{Sx_1x_2}{Sx_1^2}$ is often written $x_2 \cdot 1$</td>
</tr>
<tr>
<td>129</td>
<td>8b</td>
<td>$S$</td>
<td>$G$</td>
</tr>
<tr>
<td></td>
<td>6b</td>
<td>$S$</td>
<td>$G$</td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td></td>
<td>Add &quot;Y is unchanged&quot;</td>
</tr>
<tr>
<td>133</td>
<td>7b</td>
<td>$\pi$ equations</td>
<td>$\pi$ normal equations</td>
</tr>
<tr>
<td>139</td>
<td>12b</td>
<td></td>
<td>Add &quot;or&quot;</td>
</tr>
<tr>
<td>146</td>
<td>2b</td>
<td>$Y$</td>
<td>$\hat{Y}$</td>
</tr>
<tr>
<td>147</td>
<td>3</td>
<td>$b_{21}$</td>
<td>$b_{12}$</td>
</tr>
<tr>
<td>153</td>
<td>5</td>
<td>$S\left[S( ) \cdots \right]^2$</td>
<td>$S\left[S( ) \cdots \right]^2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Add $1/$ after &quot;composite&quot;</td>
</tr>
<tr>
<td>157</td>
<td>2</td>
<td>$R_T = R_k - \Sigma$</td>
<td>$R_T - R_k = \Sigma$</td>
</tr>
</tbody>
</table>

*9b means: 9th line from the bottom of the page.*
Errata Sheet (Con'd)

Page 158 9b  Fisher (3) and Snedecor (1)  Fisher (12) and Snedecor (10)

163  j = 1  \( P_{3j} = \frac{6}{5}, P_{3j}^t = 1 \)
    j = 5  \( P_{3j} = -\frac{6}{5}, P_{3j}^t = -1 \)

164  8  means

168  1  Snedecor (1)


184  5  Wister

193  10  Columns

201  7b  , and

205  4  \( F(.05) = 2.67 \)

211

214 Replication IV  9, 10, 11

220  2  first "r < k + 1"

222  6  effects \( t_{1j} \) and \( t_{2j} \)
    7  \( t_{1j} \)
    8  \( t_{1j} - t_{2j} \)
    11  \( \bar{t}^2 (t_{1j} - t_{2j}) \)
    15  \( \bar{t}^2 (t_{11} - t_{22}) \)

224  1, 2  2

    Change table of \( t_{ij}^t \)

235  6  \( \zeta_3 \)

237  Titlè

238  12  Ann. Math. Stat. 21:

REFERENCES CITED

<table>
<thead>
<tr>
<th>Page</th>
<th>line</th>
<th>At present</th>
<th>Should be</th>
</tr>
</thead>
<tbody>
<tr>
<td>239</td>
<td>7</td>
<td>Printice</td>
<td>Prentice</td>
</tr>
<tr>
<td></td>
<td>8b</td>
<td>Biometrika</td>
<td>Biometrika</td>
</tr>
<tr>
<td>241</td>
<td>3b</td>
<td>$\sum x_{.j}$</td>
<td>$\sum x_{.j}$</td>
</tr>
<tr>
<td>253</td>
<td>6b</td>
<td>(9)</td>
<td>(8)</td>
</tr>
<tr>
<td>256</td>
<td>11</td>
<td>Mufflin</td>
<td>Mufflin</td>
</tr>
<tr>
<td>260</td>
<td>10b</td>
<td>$\frac{\sum \sum}{\sum q \sum rq}$</td>
<td>$\frac{\sum \sum}{\sum q \sum rq}$</td>
</tr>
<tr>
<td></td>
<td>5b</td>
<td></td>
<td>Omit )</td>
</tr>
<tr>
<td></td>
<td>4b</td>
<td></td>
<td>Omit ( ,  )</td>
</tr>
<tr>
<td>279</td>
<td>5b</td>
<td>effects $(\mathbf{C} \beta)_{ij}$</td>
<td>effects, $(\mathbf{C} \beta)_{ij}$</td>
</tr>
<tr>
<td>284</td>
<td>7b</td>
<td>$\sum \sum (\varepsilon_{.i j} - \varepsilon_{.i j})$</td>
<td>$\sum \sum (\varepsilon_{.i j} + \varepsilon_{.i j})$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>main</td>
<td>main</td>
</tr>
<tr>
<td>285</td>
<td>2b</td>
<td>mean</td>
<td>$\bar{\beta} \bar{x}_i$</td>
</tr>
<tr>
<td>286</td>
<td>8b</td>
<td>$\xi_{i j}$</td>
<td>$\xi_{i j}$</td>
</tr>
<tr>
<td></td>
<td>4b</td>
<td>p</td>
<td>$\nu$</td>
</tr>
<tr>
<td>290</td>
<td>5b</td>
<td>$\theta = \frac{W_1 - W_2}{6(W_1 - W_2)}$</td>
<td>$\theta = \frac{W_1 - W_2}{6(W_1 + W_2)}$</td>
</tr>
<tr>
<td>299</td>
<td></td>
<td></td>
<td>Add to Exercise 11.3.13:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(c) Show that SSB(adj.) is independent of treatment effects.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(d) Show that $E(V_b)$ and $E(W_2)$ are as given on page 293.</td>
</tr>
<tr>
<td>300</td>
<td>4</td>
<td>$\overline{Y}_{.j}$</td>
<td>$\overline{Y}_{.j}$</td>
</tr>
<tr>
<td>302</td>
<td>10</td>
<td>$\overline{Y}_{.i}$</td>
<td>$\overline{Y}_{.i}$</td>
</tr>
<tr>
<td>303</td>
<td>1b</td>
<td>Linear</td>
<td>Linear</td>
</tr>
<tr>
<td>305</td>
<td>13</td>
<td>Monroe</td>
<td>Monroe</td>
</tr>
</tbody>
</table>
8.1 Introduction.

Up to this point, we have discussed various populations and the distributions, \( f(Y) \), of the variables or measurements, \( Y \), composing these populations. We included a discussion of the distribution of several statistics calculated from sample data taken from these populations. Among these statistics were the sample mean and variance, \( \bar{Y} \) and \( s^2 \), which were estimates of the population mean and variance, \( \mu \) and \( \sigma^2 \), respectively, where \( \mu = \text{E}(Y) \) and \( \sigma^2 = \text{E}(Y - \mu)^2 \).

We also discussed various statistics used to estimate the probability that a given sample could have come from an assumed \( \mathcal{N}(\mu, \sigma^2) \) population. These statistics included (i) the normal deviate, \( \sqrt{n} (\bar{Y} - \mu) / \sigma \), (ii) \( t = \sqrt{n} (\bar{Y} - \mu) / s \), and (iii) \( \chi^2 = (n - 1)s^2 / \sigma^2 \). We also had statistics to estimate the probability that two samples could have come from the same \( \mathcal{N}(\mu, \sigma^2) \) population: (i) \( t = (\bar{x}_1 - \bar{x}_2) / s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \), and (ii) \( F = s_1^2 / s_2^2 \).

Now we propose to consider the problem of estimating the value of some dependent variable, \( Y \), on the basis of information on one or more other variables \( X_1, X_2, \ldots \). For example, we might wish to estimate the yield of wheat, \( Y \), for different amounts, \( X \), of a standard fertilizer applied to the soil. Or more exactly, we might estimate this same yield on the basis of the amount of nitrogen, \( X_1 \), phosphate, \( X_2 \), and potash, \( X_3 \), applied to the soil. There is a great demand for information on the effect of temperature and precipitation on yields. The economist tries to predict future employment and price relationships on the basis of past data on the same and other economic variables. The engineer has the problem of estimating the probable length of life of roads or other structures in terms of such things as probable use, type of
construction and weather conditions. The doctor must decide on the basis of certain measurements how much of a given drug can be safely administered to a patient.

As indicated in a previous chapter on bivariate distributions, the expected value of \( Y \) for a single fixed \( X \) gave the so called regression curve of \( Y \) on \( X \). A first approximation to this curve was indicated to be a straight line of the form

\[
E(Y | X) \approx \alpha + \beta X.
\]

In fact if \( X \) and \( Y \) were distributed as a bivariate normal, \( E(Y | X) \equiv \alpha + \beta X \). \( Y \) is called the dependent variable and \( X \) the independent variable. However it should be understood that in general the straight line is only an approximation to the true relationship between \( Y \) and \( X \). It is well known that we can approximate a short interval of any function by a straight line; hence, if we collect data for a small range of \( X \) it is possible that a straight line would fit the data quite well even though the true relationship were curvilinear.

Let us assume that the measured value of \( Y \) can be written as

\[
Y = E(Y | X) + \epsilon,
\]

where \( \epsilon \) represents some residual or error, the amount of \( Y \) not accounted for by the regression curve of \( Y \) on \( X \). We postulate that the regression curve is selected so that the residuals are of a random nature, with the usual added assumption that the \( \epsilon \) are NID \((0, \sigma^{-2})\). If \( Y \) is a linear function of several independent variables, we might write

\[
Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \epsilon,
\]

where \( E(Y) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots \). This equation assumes that the only error involved is \( \epsilon \); in other words, there is no error in the \( X \)'s. Hence we are considering only a one-variate normal distribution with the mean of \( Y \) being approximated by a simple linear function of the \( X \)'s. If the \( X \)'s are not measured
without error (in other words the X's have probability distributions of their own),
we are led to consider the more complicated problem of multivariate analysis.
Also we are considering here only regression equations which are linear in the
regression coefficients, \( \alpha \) and the \( \beta \)'s. Methods of handling multivariate
problems and problems of non-linearity of the regression coefficients are beyond
the scope of this course. It should be emphasized that non-linearity of the X's can
be handled by the introduction of new terms in the regression equation. For
example, \( X_2 \) might very well represent \( X_1^2 \).

In order to estimate the relationship between \( Y \) and \( X \) (or between \( Y \) and
\( X_1, X_2, \ldots \) for more than one independent variable), a simultaneous observations
will be obtained on \( Y \) and \( X \). We can write each observed value \( Y_j \) in terms
of estimates of \( E(Y_j) \) and \( E(e_j) \) as

\[
Y_j = \hat{Y}_j + e_j, \quad j = 1, 2, \ldots, n,
\]

where \( \hat{Y}_j \) is the estimate of \( E(Y_j) \) and \( e_j \) the estimate of \( e_j \). Hence if
\( E(Y) \) is a linear relationship,

\[
Y = \alpha + \beta X + e = a + bX + e,
\]

where \( \alpha \approx a \) and \( \beta \approx b \). We can represent these two equations graphically,
as follows:
E(Y) is the true regression line and \( \hat{Y} \) the observed regression line. For a given observation \((X_j, Y_j)\), the true error is given by \( \varepsilon_j \), the estimated residual by \( \varepsilon_j \), and the difference between \( \hat{Y} \) and \( E(Y) \) by \( \delta_j = (\varepsilon_j - \varepsilon_j) \).

In order to obtain the values of \( a \) and \( b \), at least two sets of observations \((X, Y)\) are required. Of course if only two sets are obtained, \( \hat{Y} \) will pass through the two points, both values of \( a \) will be 0, and the values of \( a \) and \( b \) can be determined by a simultaneous solution of the two equations representing the two points. However if more than two sets of observations are obtained, we will have the situation pictured above with a sample residual, \( \varepsilon \), corresponding to each set \((X, Y)\). When there are more than two sets of points, some new method of determining \( a \) and \( b \) must be found. In the section on bivariate distributions, we indicated that \( a \) and \( b \) could be determined by the method of moments. There are many other methods of determining these estimates of the parameters, \( a \) and \( b \), in order to obtain the "best" linear fit to the data.

In any case, it seems reasonable to make the \( \{\varepsilon\} \) as small as possible. But what do we do to make these \( \{\varepsilon\} \) small? Many courses are suggested, among which are:

(i) Minimize the sum of the absolute values of the \( \varepsilon \).

(ii) Minimize the greatest of the absolute residuals.

(iii) Minimize the sum of the squares of the residuals.

It turns out that (iii) is easiest to apply, especially if we assume the true residuals \( \{\varepsilon\} \), are NID. And if the \( \{\varepsilon\} \) are NID, (iii) is also probably best from the theoretical point of view. We shall show later for NID true residuals that minimizing the sum of squared sample residuals gives the same result as the method of maximum likelihood, which is a fundamental concept in the theory of estimation. And for NID true residuals, the methods of moments and least
squares give the same results, when the regression equation is linear in the

\( Y = \beta_0 + \xi = a + \xi_0, \)

where \( SE_0^2 = S(Y - a)^2. \) It is understood that \( S \) indicates summation with respect to \( j \), the sample values. It is assumed that the true errors \( \xi \) are NID \((0, \sigma^2)\), and that \( a \) is to be determined so as to minimize \( SE_0^2 \), which we shall designate as SSE_0. The estimation equation is

\[
\frac{\partial (SSE_0)}{\partial a} = 0; \quad na = SY = n\bar{Y}.
\]

Hence \( a = \bar{Y} \). This indicates that \( \beta_0 \) which is estimated by \( a \) is the true mean of \( Y \). In terms of the true residuals,

\[ \bar{Y} = SY/n = \beta_0 + \bar{\xi}_0; \quad \xi_0 = (Y_j - \bar{Y}) = \xi_0 - \bar{\xi}_0, \]

where \( \bar{\xi}_0 = S(\xi - \bar{\xi}_0)/n \). Also

\[ SSE_0 = SE_0^2 = S(\xi_0 - \bar{\xi}_0)^2; \quad E(SSE_0) = (n - 1)\sigma^2. \]

Hence \( s_0^2 = SSE_0/(n - 1) \) is an unbiased estimate of \( \sigma^2 \). The variance of \( \bar{Y} \) is given by

\[
E(\bar{Y} - E(\bar{Y}))^2 = E(\xi_0)^2 = \sigma^2/n = s_0^2/n,
\]

since \( E(\bar{Y}) = \beta_0 \). Hence we can test the null hypothesis that \( \beta_0 = \beta_0 \)

by use of

\[ t = \sqrt{n} \frac{(\bar{Y} - \beta_0)}{s_0^2}. \]

It is evident that we are here investigating the mean of a set of data to test whether these data could have come from a normal population with mean \( \beta_0 \). Hence the method of least squares can be considered as a more
general estimation procedure with the estimation of the mean as a special case with no independent variates. It should be apparent that \( \bar{Y} \) is normally distributed, since it is a linear function of the \( \varepsilon \), which are \( \text{NID}(0, \sigma^2) \). In fact \( \bar{Y} \) is \( \text{N}(\mu, \sigma^2/n) \).

3.3 The Theory of Least Squares for \( \hat{Y} = a_1 + b_{11}X_1 \)

Now suppose that we wish to determine if the estimate of \( Y \) can be improved when we also know the value of some other variate, \( X_1 \), so that \( \hat{Y} = a_1 + b_{11}X_1 \).

The model is

\[
Y = \mu + \varepsilon_1 + b_{11}X_1 = a_1 + b_{11}X_1 + e_1.
\]

\[
\text{SSE}_1 = S(Y - a_1 - b_{11}X_1)^2.
\]

Again it is assumed that the true errors \( \{\varepsilon_1\} \) are \( \text{NID}(0, \sigma^2) \) and that the \( \{X_1\} \) are fixed values. \( a_1 \) and \( b_{11} \) are to be determined so as to minimize \( \text{SSE}_1 \).

The estimation equations are

\[
\frac{\partial (\text{SSE}_1)}{\partial a_1} = 0: \quad SY = na_1 + b_{11}SX_1
\]

\[
\frac{\partial (\text{SSE}_1)}{\partial b_{11}} = 0: \quad SX_1Y = a_1SX_1 + b_{11}SX_1^2.
\]

From the first equation we see that

\[ a_1 = \bar{Y} - b_{11}\bar{X}_1. \]

Hence \( \hat{Y} = \bar{Y} + b_{11}X_1 \), where \( X_1 = X_1 - \bar{X}_1 \). This indicates that we might simplify the problem of solving for \( b_{11} \) if we used deviations from the means \( y = Y - \bar{Y} \) and \( X_1 = X_1 - \bar{X}_1 \). Hence we might write

\[ Y = \mu + \varepsilon_1 + b_{11}X_1 = \bar{Y} + b_{11}X_1 + e_1 \text{ or } y = b_{11}x_1 + e_1, \]

The double subscript notation for \( b \) is as follows:
(a) the first applies to the number of independent variates used;
(b) the second applies to a particular one of those variates.
where \( \lambda_1 = \alpha - \beta_{11} \bar{x}_1 \). It is easily seen that

\[
b_{11} = \frac{S_{x_1}y}{S_{x_1}^2}
\]

where \( S_{x_1}y = S(x_1 - \bar{x}_1)(y - \bar{y}) = Sx_1y - (Sx_1Sy)/n \) and \( S_{x_1}^2 = Sx_1^2 - (Sx_1)^2/n \).

Also \( S_{x_1}y = Sx_1y \), since \( Sx_1 = 0 \).

If we substitute this value of \( b_{11} \) in the original equation, we find that the value of \( SSE_1 \) is

\[
SSE_1 = S(y - b_{11}x_1)^2 = S\bar{y}^2 - b_{11}Sx_1y = S\bar{y}^2 - n\bar{y}^2 - b_{11}Sx_1^2.
\]

Hence we can say that the regression has resulted in a reduction of

\[
SSR_1 = b_{11}^2S_{x_1}^2 = b_{11}Sx_1y = (Sx_1y)^2/Sx_1^2
\]

in the sum of squares for \( y \), since \( S\bar{y}^2 - n\bar{y}^2 \) was shown in Section 8.2 to have been the sum of squares of deviations from the mean. The proportional reduction in the sum of squares attributable to the regression is usually indicated by

\[
r_1^2 = (Sx_1y)^2/Sx_1^2 \cdot S\bar{y}^2.
\]

That is, \( SSE_1 = (1 - r_1^2)S\bar{y}^2 \).

In terms of the parameters, \( \alpha \) and \( \beta_{11} \), and the true errors, \( \{e_1\} \),

\[
\bar{y} = \frac{S\bar{y}}{n} = \frac{S(\alpha + \beta_{11}x_1 + e_1)}{n} = \mu + \bar{e}_1
\]

\[
y_j = y_j - \bar{y} = \beta_{11}x_{1j} + e_{1j} - \bar{e}_1
\]

\[
b_{11} = \frac{Sx_1y}{Sx_1^2} = \frac{Sx_1(\beta_{11}x_1 + e_1 - \bar{e}_1)}{Sx_1^2} = \beta_{11} + \frac{Sx_1 e_1}{Sx_1^2}
\]

\[
e_{1j} = y_j - \hat{y}_j = y_j - b_{11}x_{1j} = e_{1j} - \bar{e}_1 - \frac{Sx_1 e_1}{Sx_1^2} x_{1j}.
\]

\[
a_1 = \bar{y} - b_{11} \bar{x}_1 = \alpha_1 + (\bar{e}_1 - \frac{Sx_1 e_1}{Sx_1^2} \bar{x}_1)
\]
We have the following expectations:

\[ E(\bar{Y}) = \mu; \sigma^2(\bar{Y}) = \sigma^2/n \]

\[ E(b_{11}) = \beta_{11}; \sigma^2(b_{11}) = \sigma^2 / Sx_1^2 \]

\[ E(e_{1j}) = 0; \sigma^2(e_{1j}) = \sigma^2 \left(1 + \frac{1}{n} + \frac{x_{1j}^2}{Sx_1^2} \right) \]

\[ E(a_1) = \lambda_1; \sigma^2(a_1) = \sigma^2 \left(1 + \frac{\bar{Y}_1^2}{Sx_1^2} \right) \]

\[ \sigma^2(\bar{Y}_{11}) = \sigma^2(\bar{Y}_{e_{1j}}) = \sigma^2(b_{11}e_{1j}) = 0. \]

Hence \( \bar{Y}, b_{11} \) and \( a_1 \) are unbiased estimates of \( \mu, \beta_{11} \) and \( \lambda_1 \), respectively. Also \( \sigma^2(\bar{Y}_{e_{1j}}) = 0. \)

Regarding the error sum of squares, we see that

\[ SSE_1 = S e_1^2 = S \bar{e}_1^2 - n \bar{e}_1^2 - \frac{(Sx_1 \bar{e}_1)^2}{Sx_1^2} \]

\[ = S \bar{e}_1^2 - z_0^2 - z_1^2, \]

where \( z_0 = \sqrt{n} (\bar{Y} - \mu) \) and \( z_1 = \sqrt{Sx_1^2(b_{11} - \beta_{11})} \) are NID \((0, \sigma^2)\).

Hence the expected value of \( SSE_1 \) is \((n - 2) \sigma^2\), and \( SSE_1 \) is distributed as \( \chi^2 \sigma^2 \) with \((n - 2)\) degrees of freedom regardless of any assumption about the values of \( \mu \) and \( \beta_{11} \). Also we see that

\[ s_1^2 = SSE_1/(n - 2) \]

is an unbiased estimate of \( \sigma^2 \).

Since \( z_1^2 \) is distributed as \( \chi^2 \sigma^2 \) with \( 1 \) degree of freedom, we see that

\[ SSR_1 = b_{11}^2 Sx_1^2 = z_1^2 + \beta_{11}^2(2b_{11} - \beta_{11})Sx_1^2 \]

is also distributed as \( \chi^2 \sigma^2 \) with \( 1 \) d.f. when \( \beta_{11} = 0 \). Therefore we can use \( F = SSR_1/s_1^2 \) to test the null hypothesis that \( \beta_{11} = 0 \). We also note that
indicating that the one-tailed F-test is the appropriate test. These results are usually summarized in an analysis of variance table, as follows.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>Mean square</th>
<th>EMS*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>SSR₁</td>
<td>MSR₁ = SSR₁</td>
<td>$\sigma^2 + \beta_{11}^2 Sx_1^2$</td>
</tr>
<tr>
<td>Error</td>
<td>n - 2</td>
<td>SSE₁</td>
<td>$s_1^2$</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>

*EMS = expected value of the mean square, where the mean square is the sum of squares divided by the degrees of freedom.

Of course we could have tested $H_0: \beta_{11} = \beta_0$ by use of the t-test:

$$t = \frac{(b_{11} - \beta_0)}{\sqrt{Sx_1^2/s_1^2}}$$

$t$ having $(n - 2)$ degrees of freedom. The F-test tests $H_0: \beta_{11} = 0$; for this test, $F = t^2$.

To test the hypothesis that the Y-intercept is zero ($\beta_1 = 0$), we use

$$t = a_1/s_1 \sqrt{\frac{1}{n} + \frac{X_1^2}{Sx_1^2}}$$

If we wish to estimate the variance of the difference between a point on the computed regression line and the corresponding point on the true regression line, we see that this difference is

$$\sigma^2_{1j} = \hat{Y}_j - E(Y_j) = (\bar{Y} - \mu) + (b_{11} - \beta_{11})x_{1j},$$

with a variance of

$$\frac{1}{n} + \frac{X_1^2}{Sx_1^2} \sigma^2$$

Hence $\sigma^2_{1j}$ is a quadratic function of $x_1 (= X_1 - \bar{X}_1)$. This variance is a minimum, $\sigma^2/\bar{x}$, for $X_1 = \bar{X}_1$, and increases as $X_1$ deviates from its mean.
value. In other words, the reliability of the computed regression line in estimating the average value of Y for a future \( X_i \), \( \hat{Y}(X_i) \), decreases as the value of \( X_i \) deviates from \( \bar{X} \).

In addition the experimenter often wants to know what error to expect in predicting a particular value of \( Y_j \) for a future \( X_{1j} \) [not the average value, \( E(Y) \)], when he uses \( \hat{Y}_j \) to predict \( Y_j \). The error of prediction for a single \( Y_j \) is

\[
e_{1j} = Y_j - \hat{Y}_j.
\]

Hence the variance of this predicted value is

\[
\sigma^2(e_{1j}) = (1 + \frac{1}{n} + \frac{x_{1j}^2}{Sx^2}) \sigma^2.
\]

Or stated differently,

\[
Y_j - \hat{Y}_j = \bar{Y}_j - E(Y) - [\bar{Y}_j - E(Y)] = e_{1j} - \bar{Y}_j
\]

\[
\sigma^2(Y_j - \bar{Y}_j) = (1 + \frac{1}{n} + \frac{x_{1j}^2}{Sx^2}) \sigma^2.
\]

The confidence interval for \( \beta_{11} \) is

\[
\left[ \hat{b}_{11} - t \sigma s(\hat{b}_{11}), \hat{b}_{11} + t \sigma s(\hat{b}_{11}) \right],
\]

where \( P(t > t_{\alpha}) = \alpha/2 \). Similarly the confidence interval for the average \( Y \) for a future \( X_1 \) is

\[
(\bar{Y} + b_{11}x_1) \pm t \sigma s(\bar{Y}_1).
\]

The confidence interval is the same for a particular \( Y \) instead of the average \( Y \) except that \( s(\bar{Y}_1) \) is replaced by \( s(e_{1}) \).

Even though we have used an illustration of only one \( Y \) for each \( X \), it is understood that there was a normal population of \( Y \)'s for each \( X \) and that the residuals from the true regression line were \( \text{NID}(0, \sigma^2) \). The following assumptions were implicit in our derivations: (i) The \( X \)'s were measured without error - all of the error was in the \( Y \) direction; (ii) The variability of the \( Y \)'s about the regression line was constant from one \( X \) to another - \( \sigma^2 \) was the same for all \( X \)'s; (iii) The \( (X, Y) \) values were obtained by selecting a set of \( X \)'s and subsequently measuring the corresponding values of \( Y \). This last assumption indicates that the results cannot be applied to the entire bivariate \((X, Y)\) population - only to the set of \( X \)'s used in the analysis. C.P. Winsor has prepared an excellent discussion.
of the problem of fitting regressions when errors of measurement are present
in one or both sets of variates and when a random bivariate sample is secured (1).
Wald (2) and Bartlett (3) have presented methods of fitting a straight line when
both variables are subject to error.

If a random bivariate sample has been secured, methods have been devised to
obtain estimates of and confidence limits for the value of \( X \) for a future \( Y \) as
well as for the value of \( Y \) for a future \( X \); these will be discussed in Part III
of this text. The method of least squares, regarding \( X \) as fixed, produces the
same results as the bivariate solution when predicting \( Y \) for a future \( X \) but not
for the inverse problem of predicting \( X \) for a future \( Y \). Eisenhart (4), Bliss (5)
and Winsor (1) have discussed the latter problem when least squares have been
used to estimate the regression.

As an example of the use of a single independent variate, we might consider
the problem of estimating the quantity of Vitamin B\(_2\) in mg/gm. for turnip
greens (\( Y \)) from a knowledge of the radiation in relative gram calories/cm\(^2\)/min.
during the preceding \( \frac{1}{2} \) day of sunlight (\( X_1 \)). 27 sets of observations on these
two variables were taken. In addition observations were taken on the average soil
moisture tension (\( X_2 \)), air temperature in degrees Fahrenheit (\( X_3 \)), and the product
of \( X_1 \) and \( X_2 \) (\( X_4 \)). These other independent variables will be considered in subsequent
sections but are included in the table below to save space. In this table, \( X_4 \)
has been divided by 10.

These data were secured in experiments conducted by the Southern Cooperative
Group.
Using only $X_1$ as a predictor, the regression equation is

$$Y = \hat{\alpha} + \hat{\beta}X_1 + \varepsilon_1 = \hat{Y} + \varepsilon_1,$$

where $\hat{Y} = \overline{Y} + \hat{b}_{11}X_1$. The computations are as follows:

$$\begin{align*}
n &= 27 \quad SY = 2273.5 \quad SX_1 = 5016 \\
Sx_1Y &= SX_1Y - (SX_1)(SY)/n = 425,532.3 - 422,365.8 = 3,166.5 \\
Sx_1^2 &= SX_1^2 - (SX_1)^2/n = 1,034,438 - 931,861.33 = 102,576.67 \\
Sy^2 &= SY^2 - (SY)^2/n = 200,587.65 - 191,437.12 = 9,150.53 \\
b_{11} &= \frac{Sx_1Y}{Sx_1^2} = \frac{3,166.5}{102,576.67} = 0.0308696
\end{align*}$$
\[ \text{SSR}_1 = b_{11} S_{x_1 y} = \frac{(S_{x_1 y})^2}{S_{x_1}^2} = 97.75 \]

\[ \text{SSE}_1 = \text{Sy}^2 - \text{SSR}_1 = 9.150.53 - 97.75 = 9.052.78 \]

\[ s_1^2 = \frac{\text{SSE}_1}{25} = 362.11 \]

\[ s(b_{11}) = \sqrt{\frac{s_1^2}{S_{x_1}^2}} = \sqrt{362.11/102.577} = 0.059415 \]

The analysis of variance table is

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>Mean squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>97.75</td>
<td>97.75</td>
</tr>
<tr>
<td>Error</td>
<td>25</td>
<td>9.052.78</td>
<td>362.11</td>
</tr>
</tbody>
</table>

To test the null hypothesis \( H_0: \beta_{11} = 0 \), we consider \( F = \frac{97.75}{362.11} = 0.27 \).

Since \( F < 1 \), we do not reject the null hypothesis. In other words, these data do not furnish any evidence to support the theory that Vitamin B2 content can be estimated from radiation. We also note that for the same \( H_0 \),

\[ t = \frac{0.03087/0.0594}{0.52} = \sqrt{F} = 0.52 \]

In this case, there is no need to compare the predicted values on the regression line with the true values, because we have shown that the regression line did not improve the prediction over the use of a simple \( Y \approx \bar{Y} \).

Exercise 8.3.1: Use the confidence limits for predicting the average \( Y \) for a future \( X_1 \) to show that the confidence limits of \( X_1 \) for a future known \( Y_0 \) are

\[ X_1 = \bar{X}_1 + b_{11}(Y_0 - \bar{Y}) + t_{*} \frac{s_1}{\lambda} \sqrt{\frac{1}{n} + \frac{(Y_0 - \bar{Y})^2}{S_{x_1}^2}} \]

where

\[ \lambda = b_{11}^2 - \frac{t_{*}^2 s_1^2}{S_{x_1}^2} \]

Exercise 8.3.2: What changes would be made in the confidence limits for \( X_1 \) if \( Y_0 \) was the average of \( k \) observations?

Exercise 8.3.1. if \( Y_0 \) was the average of \( k \) observations?
Exercise 8.3.3.

A study was made of the relationship between the net income (Y) of southern farm families and a socio-economic score (X), the latter based on the possession of certain items such as radio, telephone, automobile, and electricity and the education and church attendance of the heads of the families (6). The possible range of X's was 39-91.

(a) Set up a linear regression equation to predict income from socio-economic score.

(b) Given the following data, compute estimates of the parameters in (a) and their standard errors, set up the analysis of variance (Y is in thousands of dollars).

\[ n = 909 \quad SY = 2,225 \quad SX = 51,852 \]
\[ SX^2 = 3,055,833 \quad SXY = 133,473 \quad SY^2 = 7,439 \]

(c) Is the regression coefficient significantly different from zero? What are its 95% confidence limits?

(d) What percentage of the total variability is accounted for by the regression?

(e) Estimate the true net income for \( X = 80 \). Compute the 95% confidence limits for this true income. Within what limits, with 95% confidence, would you expect a particular family's net income for \( X = 80 \)?

(f) Since it is much easier to obtain the data upon which the socio-economic score is based instead of income data, one would prefer using the score instead of income to classify families. Do you think that the score is a suitable replacement for income in classifying these southern farm families? Why?

Exercise 8.3.4.

The analysis of scores is simplified for those items with only two alternatives, e.g., with or without electricity. Suppose we want to correlate the scores on one such item with income.

Let \( w_0 \) be the score for each of the \( n_0 \) families without this item and \( w_1 \) the score for each of the \( n_1 \) families with this item (\( n_0 + n_1 = n \) total families).
Show that \( r^2 \) is independent of \( w_0 \) and \( w_1 \).

**Exercise 8.3.5.**

Girshick and Haavelmo have made an analysis of the demand for food in the United States for the years 1922-1941 (7). One equation in their analysis involved the relationship between disposable income adjusted for the cost of living (\( Y \)) and investment per capita adjusted for the cost of living (\( X_1 \)). The values of \( Y \) and \( X_1 \) are

<table>
<thead>
<tr>
<th>( Y )</th>
<th>( X_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>87.4</td>
<td>92.9</td>
</tr>
<tr>
<td>97.6</td>
<td>142.9</td>
</tr>
<tr>
<td>96.7</td>
<td>100.0</td>
</tr>
<tr>
<td>98.2</td>
<td>123.8</td>
</tr>
<tr>
<td>99.8</td>
<td>111.9</td>
</tr>
<tr>
<td>100.5</td>
<td>121.4</td>
</tr>
<tr>
<td>103.2</td>
<td>107.1</td>
</tr>
<tr>
<td>107.8</td>
<td>142.9</td>
</tr>
<tr>
<td>96.6</td>
<td>92.9</td>
</tr>
<tr>
<td>88.9</td>
<td>97.6</td>
</tr>
<tr>
<td>75.1</td>
<td>52.4</td>
</tr>
<tr>
<td>76.9</td>
<td>40.5</td>
</tr>
<tr>
<td>84.6</td>
<td>64.3</td>
</tr>
<tr>
<td>90.6</td>
<td>78.6</td>
</tr>
<tr>
<td>103.1</td>
<td>114.3</td>
</tr>
<tr>
<td>105.1</td>
<td>121.4</td>
</tr>
<tr>
<td>96.4</td>
<td>78.6</td>
</tr>
<tr>
<td>104.4</td>
<td>109.5</td>
</tr>
<tr>
<td>110.7</td>
<td>128.6</td>
</tr>
<tr>
<td>( \frac{127.1}{1950.7} )</td>
<td>( \frac{238.1}{2159.7} )</td>
</tr>
</tbody>
</table>

(a) Set up a simple linear regression of \( Y \) or \( X_1 \) and determine the constants in the regression equation.

(b) Set up the analysis of variance.

(c) Make a test of significance of the usefulness of the regression equation. Are there any aspects of these data which might invalidate this test?

(d) Plot the data (rounded to nearest integer) and draw in the regression line and the 95% confidence lines for \( E(Y|X_1) \). From the nature of the residuals from the regression line, would you suggest any changes in the form of the regression equation?
Exercise 8.3.6.

R. A. Fisher has compared the body weights (in kilograms) with the heart weights (in grams) of 47 female and 97 male cats (8). The sums of squares and products were as follows:

<table>
<thead>
<tr>
<th></th>
<th>Females</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D.F.</td>
<td>(Body)²</td>
<td>(Body • Heart)</td>
</tr>
<tr>
<td>Total</td>
<td>47</td>
<td>265.13</td>
<td>1029.62</td>
</tr>
<tr>
<td>Correction for mean</td>
<td>1</td>
<td>261.677</td>
<td>1020.516</td>
</tr>
<tr>
<td>Difference</td>
<td>46</td>
<td>3.453</td>
<td>9.104</td>
</tr>
<tr>
<td></td>
<td>Males</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>97</td>
<td>836.75</td>
<td>3275.55</td>
</tr>
<tr>
<td>Correction for mean</td>
<td>1</td>
<td>815.77</td>
<td>3185.07</td>
</tr>
<tr>
<td>Difference</td>
<td>96</td>
<td>20.98</td>
<td>90.48</td>
</tr>
</tbody>
</table>

(a) Determine the regression of heart weight on body weight for both males and females.

(b) Are these two regressions different from one another?

(c) Are the two error variances essentially the same?

Exercise 8.3.7.

In a study of lobster population, D. B. DeLury presents the following data on the catch per unit of effort for the time interval t, C(t), and the total catch up to t, K(t), in thousands of pounds.

<table>
<thead>
<tr>
<th>t</th>
<th>C</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.82</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>.75</td>
<td>.7</td>
</tr>
<tr>
<td>3</td>
<td>.94</td>
<td>.13</td>
</tr>
<tr>
<td>4</td>
<td>.80</td>
<td>.16</td>
</tr>
<tr>
<td>5</td>
<td>.83</td>
<td>.22</td>
</tr>
<tr>
<td>6</td>
<td>.99</td>
<td>.25</td>
</tr>
<tr>
<td>7</td>
<td>.70</td>
<td>.32</td>
</tr>
<tr>
<td>8</td>
<td>.58</td>
<td>.37</td>
</tr>
<tr>
<td>9</td>
<td>.64</td>
<td>.40</td>
</tr>
<tr>
<td>10</td>
<td>.55</td>
<td>.45</td>
</tr>
<tr>
<td>11</td>
<td>.52</td>
<td>.49</td>
</tr>
<tr>
<td>12</td>
<td>.45</td>
<td>.45</td>
</tr>
<tr>
<td>13</td>
<td>.48</td>
<td>.48</td>
</tr>
<tr>
<td>14</td>
<td>.43</td>
<td></td>
</tr>
</tbody>
</table>

(a) A linear equation of the form \( C = a + bk + e \) was set up. Determine the values of a and b and their standard errors.

(b) The total population at time \( t = 0 \) is estimated by \( N_0 = -a/b \). Determine \( N_0 \).
Exercise 8.3.8.

Read a brief note in the December, 1948 issue of the American Statistician, pages 16-17, on the use of regression methods for business statistics.

Exercise 8.3.9.

Suppose a sample of $n_1$ is used to estimate the parameters in the equation

$$y_1 = \mu_1 + \beta_1 x_1 + \epsilon_1$$

and a sample of $n_2$ for the equation

$$y_2 = \mu_2 + \beta_2 x_2 + \epsilon_2,$$

where $\sigma_1^2$ is assumed to equal $\sigma_2^2$. How would you test the null hypothesis that $\beta_1 = \beta_2$?
8.4, The Theory of Least Squares for $\hat{Y} = \bar{Y} + b_{21}x_1 + b_{22}x_2$.

It may be possible to improve the prediction of $Y$ by adding a second independent variable, $x_2$. Or it may be that we have theoretical reasons to believe that

$$E(Y) = \mu + \beta_{21}x_1 + \beta_{22}x_2.$$

That is $Y = \mu + \beta_{21}x_1 + \beta_{22}x_2 + \epsilon$. Note that we denote the true regression coefficient for $x_1$ as $\beta_{21}$ instead of $\beta_{11}$. We will see later that the two regression coefficients, $\beta_{11}$ and $\beta_{21}$, are truly different.

In this case we have the following equations to determine $b_{21}$ and $b_{22}$, which are called the normal equations:

$$\begin{cases} b_{21}Sx_1^2 + b_{22}Sx_1x_2 = Sx_1y \\ b_{21}Sx_1x_2 + b_{22}Sx_2^2 = Sx_2y \end{cases}.$$

The solutions are

$$\begin{cases} b_{21} = c_{11}Sx_1y + c_{12}Sx_2y \\ b_{22} = c_{21}Sx_1y + c_{22}Sx_2y \end{cases},$$

where $c_{11} = Sx_2^2/DSx_1^2$, $c_{12} = c_{21} = -Sx_1x_2/DSx_1^2$, $c_{22} = 1/D$, and $D = Sx_2^2 - (Sx_1x_2)^2/Sx_1^2$.

For those students who are acquainted with matrix algebra methods, we express the two normal equations and the solutions as

$$AB = S; \quad B = A^{-1}S = CS,$$

where

$$A = \begin{bmatrix} Sx_1^2 (=c_{11}) & Sx_1x_2 (=c_{12}) \\ Sx_1x_2 (=c_{21}) & Sx_2^2 (=c_{22}) \end{bmatrix}, \quad B = \begin{bmatrix} b_{21} \\ b_{22} \end{bmatrix}, \quad S = \begin{bmatrix} Sx_1y (=s_1) \\ Sx_2y (=s_2) \end{bmatrix}.$$
of the cross-products with the dependent variable, and \( C \) the inverse of the
original A matrix. The \( c_{ij} \) satisfy these relations:
\[
\sum_j c_{ij} e_i e_j = \begin{cases} 0 & i \neq i' \\ 1 & i = i' \end{cases}
\]
That is, the sum of products of the same row is 1 and of unlike rows is 0.

Returning to the solutions for \( b_{21} \) and \( b_{22} \) given above, we can express these
solutions in terms of the parameters and \( \varepsilon \), in the original equation as follows:

\[
\bar{y} = \mu + \bar{e} ; \quad E(\bar{y}) = \mu
\]
\[
y = y - \bar{y} = \beta_{21} x_1 + \beta_{22} x_2 + \varepsilon
\]

\[
b_{21} = c_{11} sx_1 y + c_{12} sx_2 y = \beta_{21} (c_{11} sx_1^2 + c_{12} sx_1 x_2)
\]
\[
+ \beta_{22} (c_{11} sx_1 x_2 + c_{12} sx_2^2) + c_{11} sx_1 \varepsilon + c_{12} sx_2 \varepsilon
\]
\[
= \beta_{21} + c_{11} sx_1 \varepsilon + c_{12} sx_2 \varepsilon
\]

\[
E(b_{21}) = \beta_{21}
\]

\[
\sigma^2(b_{21}) = \left[c_{11}(c_{11} sx_1^2 + c_{12} sx_1 x_2) + c_{12}(c_{11} sx_1 x_2 + c_{12} sx_2^2)\right] \sigma^2
\]
\[
= c_{11} \sigma^2
\]

Similarly \( b_{22} = \beta_{22} + c_{21} sx_1 \varepsilon + c_{22} sx_2 \varepsilon \), \( E(b_{22}) = \beta_{22} \) and

\[
\sigma^2(b_{22}) = c_{22} \sigma^2.
\]

\[
\sigma(b_{21}, b_{22}) = \left[c_{11}(c_{21} sx_1^2 + c_{22} sx_1 x_2) + c_{12}(c_{21} sx_1 x_2 + c_{22} sx_2^2)\right] \sigma^2
\]
\[
= c_{12} \sigma^2
\]

\[
\gamma(\bar{y}, b_{21}) = \gamma(\bar{y}, b_{22}) = 0. \text{ Hence } \bar{y}, b_{21} \text{ and } b_{22} \text{ follow a multivariate normal}
\]
distribution with variances \( \sigma^2/n, c_{11} \sigma^2, c_{22} \sigma^2 \) and covariance 0 and \( c_{12} \sigma^2 \).

The error sum of squares is

\[
SSE_2 = S(y - b_{21} x_1 - b_{22} x_2)^2 = Sy^2 - b_{21} Sx_1 y - b_{22} Sx_2 y + b_{21} (b_{21} Sx_1^2 + b_{22} Sx_1 x_2)
\]
\[
- Sx_1 y) + b_{22} (b_{22} Sx_2^2 + b_{21} Sx_1 x_2 - Sx_2 y) = Sy^2 - b_{21} Sx_1 y - b_{22} Sx_2 y.
\]

In order to study the distribution of \( SSR_2 \), we first set up two estimation
equations:

\[
y = b_{11} x_1 + e_1 \quad \text{and} \quad x_2 = b_{21} x_1 + e_x
\]
from which we determine by least squares that

\[ b_{11} = \frac{Sx_1 y}{Sx_1^2} \]  and \[ b_x = \frac{Sx_1 x_2}{Sx_1^2}. \]

Then we seek to reduce \( Se_x^2 \) by the following equation

\[ e_1 = b_{22} e_x + e_2 = b_{22}(x_2 - bx_1) + e_2, \]

where we determine by minimizing \( Se_x^2 \) that

\[ b_{22} = \frac{Se_x e_x}{Se_x^2} = \frac{Sy - b_{11} x_1}{Se_x^2} = \frac{Sx_1 e_x}{Se_x^2}. \]

since \( Sx_1 e_x = Sx_1 x_2 - b_x Sx_1^2 = 0 \). Hence

\[ y = b_{11} x_1 + e_1 = (b_{11} - b_{22} \frac{Sx_1 x_2}{Sx_1^2}) x_1 + b_{22} x_2 + e_2 \]

\[ = b_{11} x_1 + b_{22}(x_2 - \frac{b_{22} Sx_1 x_2}{Sx_1^2}) + e_2. \]

Since

\[ b_{21} = \frac{Sx_1 y - b_{22} Sx_1 x_2}{Sx_1^2} = b_{11} - b_{22} \frac{Sx_1 x_2}{Sx_1^2}, \]

we see that this last approach to the regression problem produces the same result as given at the beginning of this section 8.4. But we note that, using the last part of the above equation, the regression of \( y \) on \( x_1 \) and \( x_2 \) can be regarded as the regression on \( x_1 \) and \( x_2 \) adjusted for \( x_1 \). The total reduction is the sum of two independent reductions

\[ SSR_2 = b_{11} Sx_1 y + b_{22} Se_x e_x = b_{11} Sx_1 y + b_{22} Sx_e y = \frac{(Sx_1 y)^2}{Sx_1^2} + \frac{(Se_x y)^2}{Se_x^2}, \]

the reduction due to \( x_1 \) alone (section 8.3) plus the added reduction due to \( x_2 \). The latter is in terms of the reduction due to \( x_2 \) adjusted for \( x_1 \) (\( e_x = x_2 - bx_1 \)).

Using the model given at the beginning of this section 8.4, we see that the reduction due to fitting \( x_1 \) alone and the expected value of this reduction are
\[ z_1^2 = \frac{(Sx_1y)^2}{Sx_1^2} = \left( \frac{\beta_{21}Sx_1^2 + \beta_{22}Sx_1x_2 + Sx_1\varepsilon}{Sx_1^2} \right)^2 \]

\[ E(z_1^2) = (\beta_{21} + \beta_{22} \frac{Sx_1x_2}{Sx_1^2})^2 Sx_1^2 + \sigma^2 \]

Similarly the added reduction due to \( x_2 \) (after \( x_1 \)) and its expected value are:

\[ z_2^2 = \frac{\left[ S(x_2 - b_x)x_1 \right]^2}{D} = \frac{\left[ D\beta_{22} + S\varepsilon (x_2 - b_xx_1) \right]^2}{D} \]

\[ E(z_2^2) = D \beta_{22}^2 + \sigma^2 \]

where

\[ b_x = \frac{Sx_1x_2}{Sx_1^2} \quad \text{and} \quad D = Sx_2^2 - \frac{(Sx_1x_2)^2}{Sx_1^2} \]

It should be noted that \( z_2^2 \) can be most easily computed as \( SSR_2 = z_2^2 \), where

\[ SSR_2 = \beta_{21}Sx_1y + \beta_{22}Sx_2y \]

The expected value of the total reduction, \( SSR_2 = z_1^2 + z_2^2 \), is

\[ E(SSR_2) = \beta_{21}^2 Sx_1^2 + \beta_{22}^2 Sx_2^2 + 2 \beta_{21} \beta_{22} Sx_1x_2 + 2 \sigma^2 \]

Hence the error sum of squares and its expected value are:

\[ SSE_2 = Sy^2 - SSR_2 = S(\beta_{21}x_1 + \beta_{22}x_2 + \varepsilon - \bar{\varepsilon})^2 - SSR_2 \]

\[ E(SSE_2) = (n - 1)\sigma^2 - 2\sigma^2 = (n - 3)\sigma^2 \]

An unbiased estimate of \( \sigma^2 \) is given by

\[ s_2^2 = \frac{SSE_2}{n - 3} \]

Since the estimated variance of \( \beta_{21} \) is \( \sigma_{21}^2 \) and of \( \beta_{22} \) is \( \sigma_{22}^2 \), we can use the t-test to test for the significance of each regression coefficient. The proportional reduction in sum of squares due to regression is

\[ R^2 = SSR_2 / Sy^2 \]

R is called the multiple correlation coefficient.
A test of the composite hypothesis, $H_0: \beta_{21} = \beta_{22} = 0$, can be made by use of
\[
F = \frac{s_2^2 + s_2^2}{2s_2^2} = \frac{MSR_2}{s_2^2} = \frac{R^2}{2}(1 - R^2) / \frac{n - 3}{n - 3}
\]
with $2$ and $n - 3$ degrees of freedom. Considering the expectation of $z_2^2$, we see that it is possible to test $H_0: \beta_{22} = 0$ by use of $F = z_2^2 / s_2^2$, with $1$ and $n - 3$ degrees of freedom. Of course this is the same as $t^2 = b_{22}^2 / c_{22}s_2^2$. Hence we can test the usefulness of $x_2$ as a predictor without determining the inverse matrix.

If it turned out that $x_2$ was an important predictor, the experimenter might wish to know if it alone might do the job; that is, to test if $x_1$ would add any information in addition to that supplied by $x_2$. But we have no test of this except the $t$-test since $z_2^2$ cannot be used to test the single hypothesis $\beta_{21} = 0$ ($Ez_1^2$ is a function of both $\beta_{21}$ and $\beta_{22}$). However we can meet this difficulty by considering the regression of $y$ on $x_2$ and $x_1$ adjusted for $x_2$ so that SSR$_2$ is composed of the reduction due to $x_2$ alone plus the added reduction due to $x_1$. In this case we have
\[
z_1^2 = \frac{(Sx_2y)^2}{Sx_2^2}; \quad E(z_1^2) = (\beta_{22} + \beta_{21} \frac{Sx_1x_2}{Sx_2^2})^2 Sx_2^2 + \sigma^2
\]
\[
z_2^2 = \frac{[S(x_1 - b'_x x_2)y]^2}{S(x_1 - b'_x x_2)^2}; \quad E(z_2^2) = D' \beta_{21}^2 + \sigma^2,
\]
where $b'_x = \frac{Sx_1x_2}{Sx_2^2}$ and $D' = Sx_2^2 - \frac{(Sx_1x_2)^2}{Sx_2^2}$. As before $z_2^2$ is most easily computed as SSR$_2 - z_1^2$. Hence we can test $H_0: \beta_{21} = 0$ by use of
\[
F = \frac{z_2^2}{s_2^2}
\]
with $1$ and $n - 3$ degrees of freedom.

These results can be summarized in an analysis of variance table, as follows:
<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Mean square</th>
<th>EMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression on $x_1$</td>
<td>1</td>
<td>$Z_1^2$</td>
<td>$(\beta_{21} + \beta_{22}\frac{Sx_1x_2}{Sx_1^2})^2 Sx_1 + \sigma^2$</td>
</tr>
<tr>
<td>Added reduction due to $x_2$</td>
<td>1</td>
<td>$Z_2^2$</td>
<td>$D\beta_{22} + \sigma^2$</td>
</tr>
<tr>
<td>Regression on $x_2$</td>
<td>1</td>
<td>$Z_1^2$</td>
<td>$(\beta_{22} + \beta_{21}\frac{Sx_1x_2}{Sx_2^2})^2 Sx_2 + \sigma^2$</td>
</tr>
<tr>
<td>Added reduction due to $x_1$</td>
<td>1</td>
<td>$Z_2^2$</td>
<td>$D'\beta_{21} + \sigma^2$</td>
</tr>
<tr>
<td>Error</td>
<td>$n - 3$</td>
<td>$e_2^2$</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>

When two independent variables are used for prediction purposes, we have

$$
S_{2j} = (\bar{y} - \mu) + (b_{21} - \beta_{21})x_{1j} + (b_{22} - \beta_{22})x_{2j}
$$

$$
\sigma^2(S_{2j}) = \sigma^2\left(\frac{1}{n} + c_{11}x_{1j}^2 + c_{22}x_{2j}^2 + 2c_{12}x_{1j}x_{2j}\right)
$$

$$
e_{2j} = y_{ij} - S_{2j}
$$

$$
\sigma^2(e_{2j}) = \sigma^2 + \sigma^2(S_{2j})
$$

As an example of the use of two independent variables, suppose we consider the Vitamin B$_2$ problem with both $X_1$ and $X_2$ in the regression equation:

$$
Y = \mu + \beta_{21}x_1 + \beta_{22}x_2 + \epsilon
$$

The added computations required are:

$Sx_2 = 507.6$

$Sx_2^2 = 20,697.84 - 9,542.88 = 11,154.96$

$Sx_1x_2 = 94,630.6 - 94,300.8 = 379.8$

$Sx_2y = 34,058.96 - 42,741.80 = -8,683.74$

The normal equations and solutions are:

$$
102,576.67 b_{21} + 379.80 b_{22} = 1.166.50
$$

$$
379.80 b_{21} + 11,154.96 b_{22} = -8,683.74
$$

$$
D = 11,154.96 - 1.41 = 11,153.55
$$
\[ c_{22} = 0.00008965755 \quad c_{11} = 0.00000975004 \]
\[ c_{12} = c_{21} = -0.000000331966 \]

\[ b_{21} = c_{11}S_{x1y} + c_{12}S_{x2y} = 0.0337562 \]
\[ b_{22} = c_{12}S_{x1y} + c_{22}S_{x2y} = -0.779614 \]
\[ b_{12} = S_{x2y}/S_{x2}^2 = -0.778464 \]

\[ SSR_2 = b_{21}S_{x1y} + b_{22}S_{x2y} = 6,876.85 \]
\[ SSE_2 = S_y^2 - SSR_2 = 9,150.53 - 6,876.85 = 2,273.68 \]
\[ s^2 = 2,273.68/24 = 94.74 \]
\[ s(b_{21}) = \sqrt{94.74c_{11}} = 0.03039 \quad t_1 = 1.11 \]
\[ s(b_{22}) = \sqrt{94.74c_{22}} = 0.09216 \quad t_2 = 8.46 \quad (P < .01) \]

Hence we conclude that \( b_{22} \) is significant but that \( b_{21} \) is not significant; that is, we reject the null hypothesis that \( \beta_2 = 0 \) but not the hypothesis that \( \beta_1 = 0 \).

The analysis of variance is

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>Mean square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>2</td>
<td>6,876.85</td>
<td>3,438.42</td>
</tr>
<tr>
<td>( X_1 )</td>
<td>1</td>
<td>97.75</td>
<td>97.75</td>
</tr>
<tr>
<td>( X_2 ) after ( X_1 )</td>
<td>1</td>
<td>6,779.10</td>
<td>6,779.10**</td>
</tr>
<tr>
<td>( X_2 )</td>
<td>1</td>
<td>6,759.98</td>
<td>6,759.98</td>
</tr>
<tr>
<td>( X_1 ) after ( X_2 )</td>
<td>1</td>
<td>116.87</td>
<td>116.87</td>
</tr>
<tr>
<td>Error</td>
<td>24</td>
<td>2,273.68</td>
<td>94.74</td>
</tr>
</tbody>
</table>

The test for the usefulness of \( X_2 \) is

\[ F_2 = 6,779.10/94.74 = 71.55 = t_2^2 \]

and for \( X_1 \)

\[ F_1 = 116.87/94.74 = 1.23 = t_1^2 . \]

The symbol ** indicates that the probability of obtaining this on a larger mean square when the null hypothesis is true (\( \beta_2 = 0 \)) is less than 0.01. We see that \( R^2 = 6,877/9,151 = .752 \), whereas \( r^2 = 6,760/9,151 = .739 \), only slightly smaller.
On the basis of this sample, we would infer that average soil moisture \( (X_2) \) is an important factor in determining the Vitamin B\(_2\) content of turnip greens and that radiation \( (X_1) \) is relatively unimportant. In fact if we use only \( X_2 \) in the prediction equation, we have

\[
Y = \bar{Y} - 0.778464x_2
\]

with \( s^2 = (2273.68 + 116.87)/25 = 95.62 \). Hence the estimate of the error variance is not materially increased if we omit \( X_1 \) from the regression equation. In this case the standard error of the \( X_2 \) regression coefficient \( (b_{12}) \) is

\[
\sqrt{95.62/5x_2^2} = \sqrt{0.008572} = 0.09259
\]

and \( t = 0.778464/0.09259 = 8.41 \).

It should be emphasized that if only \( X_1 \) is used as a predictor the estimate of random error is biased upwards because this estimate includes the effect of \( X_2 (s_1^2 = 362 \) and \( s_2^2 = 95) \). This does not assure us that \( X_2 \) is the only useful variable to predict \( Y \), but at least no other variable can reduce the error variance materially below that obtained by using \( X_2 \). In the following sections, we shall investigate the use of the other variables, \( X_3 \) and \( X_4 \), given in the table on page 116.

One point in computing techniques should be mentioned before we consider more than two independent variables. The reader will note that the \( c_{ij} \) values were very small and that 12 decimal places had to be used for \( c_{12} \). This was a result of the large values of \( a_{ij} \). In general it is advisable to code the original data so that the \( a_{ij} \) are reduced to values between 1 and 10, if possible. In this Vitamin B\(_2\) example, we could have simplified the computations by dividing \( X_1 \) and \( X_2 \) each by 100. In this case the \( A \) and \( S \) matrices would have been

\[
A = \begin{bmatrix} 10.257667 & 0.037980 \\ 0.037980 & 1.115496 \end{bmatrix}, \quad S = \begin{bmatrix} 31.6650 \\ -86.8374 \end{bmatrix}
\]

and the \( C \) matrix

\[
\begin{bmatrix} 0.0975004 & -0.00331965 \\ -0.00331965 & 0.8965755 \end{bmatrix}.
\]

Hence \( b_{21}^1 = 3.37562 \) and \( b_{22}^1 = -77.9614 \), 100 times as large as the former \( b \)'s. But since the \( X \)'s have been divided by 100, the effect on \( Y \) is unchanged.
Exercise 8.4.1. Given \( Y = \mu + \beta_{21}x_1 + \beta_{22}x_2 + \epsilon \), but we estimate only with \( \hat{Y} = \bar{y} + b_{11}x_1 \). Show that \( b_{11} \) is a biased estimate of \( \beta_{21} \) and that \( s_1^2 \) is also a biased estimate of \( \sigma^2 \).

Exercise 8.4.2. Given \( Y = \mu + \beta_{11}x_1 + \epsilon \) but we estimate with \( \hat{Y} = \bar{y} + b_{11}x_1 + b_{22}x_2 \). Determine if \( b_{21} \) and \( s_2^2 \) are unbiased estimates of \( \beta_{11} \) and \( \sigma^2 \), respectively.

Exercise 8.4.3. How would you test the hypothesis that \( \beta_{21} = \beta_{22} \)?

Exercise 8.4.4. Plot the \( Y \) and \( X_2 \) data for the Vitamin B_2 example and draw in the regression line of \( Y \) on \( X_2 \) and the confidence limits for \( E(Y \mid X_2) \).

Exercise 8.4.5. Set up the regression equation \( \hat{Y} = a + b_1x_1 + b_2x_1^2 \). Is the added reduction due to \( b_2 \) significant?

Exercise 8.4.6. Prove that \( \text{MSR}_2 / s_2^2 \) is distributed as \( F \) with 2 and \( (n - 3) \) degrees of freedom.

Exercise 8.4.7. In the analysis of the data given in Exercise 8.3.5, it was thought advisable to use as a second independent variate \( (X_2) \) the value of \( Y \) for the previous year - 77.4, 87.4, 97.6, ..., 110.7.

(a) Estimate the two regression coefficients and their standard errors.

(b) Set up the complete analysis of variance.

(c) Estimate the average \( Y \) for \( X = 100 \) and \( X_2 = 100 \), and the standard error of this estimate.

Exercise 8.4.8. W. G. Woltz, Department of Agronomy, and W. D. Foster, Institute of Statistics, North Carolina State College, Raleigh, in 1949 analyzed 25 samples of tobacco leaf for organic and inorganic chemical constituents, and multiple regression was used to discover the nature and extent of the relationship of certain of these constituents. The dependent variables considered were rate of cigarette burn in inches per 1000 seconds \( (Y_1) \), percent sugar in the leaf \( (Y_2) \) and percent nicotine \( (Y_3) \). The independent variables were percentages of total nitrogen \( (X_1) \),
of chlorine \((X_2)\), of potassium \((X_3)\), of phosphorus \((X_4)\), of calcium \((X_5)\) and of magnesium \((X_6)\). The original data (25 observations) and corrected sums of squares and products are given below.

<table>
<thead>
<tr>
<th>(Y_1)</th>
<th>(Y_2)</th>
<th>(Y_3)</th>
<th>(X_1)</th>
<th>(X_2)</th>
<th>(X_3)</th>
<th>(X_4)</th>
<th>(X_5)</th>
<th>(X_6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.55</td>
<td>20.05</td>
<td>1.38</td>
<td>2.02</td>
<td>2.90</td>
<td>2.17</td>
<td>.51</td>
<td>3.47</td>
<td>.91</td>
</tr>
<tr>
<td>1.63</td>
<td>12.58</td>
<td>2.64</td>
<td>2.62</td>
<td>2.78</td>
<td>1.72</td>
<td>.50</td>
<td>4.57</td>
<td>1.25</td>
</tr>
<tr>
<td>1.66</td>
<td>18.56</td>
<td>1.56</td>
<td>2.08</td>
<td>2.68</td>
<td>2.40</td>
<td>.43</td>
<td>3.52</td>
<td>.82</td>
</tr>
<tr>
<td>1.52</td>
<td>18.56</td>
<td>2.22</td>
<td>2.20</td>
<td>3.17</td>
<td>2.06</td>
<td>.52</td>
<td>3.69</td>
<td>.97</td>
</tr>
<tr>
<td>1.70</td>
<td>14.02</td>
<td>2.85</td>
<td>2.38</td>
<td>2.52</td>
<td>2.18</td>
<td>.42</td>
<td>4.01</td>
<td>1.12</td>
</tr>
<tr>
<td>1.68</td>
<td>15.64</td>
<td>1.24</td>
<td>2.03</td>
<td>2.56</td>
<td>2.57</td>
<td>.44</td>
<td>2.79</td>
<td>.82</td>
</tr>
<tr>
<td>1.78</td>
<td>14.52</td>
<td>2.86</td>
<td>2.87</td>
<td>2.67</td>
<td>2.64</td>
<td>.50</td>
<td>3.92</td>
<td>1.06</td>
</tr>
<tr>
<td>1.57</td>
<td>18.52</td>
<td>2.18</td>
<td>1.88</td>
<td>2.58</td>
<td>2.22</td>
<td>.49</td>
<td>3.58</td>
<td>1.01</td>
</tr>
<tr>
<td>1.60</td>
<td>17.84</td>
<td>1.65</td>
<td>1.93</td>
<td>2.26</td>
<td>2.15</td>
<td>.56</td>
<td>3.57</td>
<td>.92</td>
</tr>
<tr>
<td>1.52</td>
<td>13.38</td>
<td>3.28</td>
<td>2.57</td>
<td>1.74</td>
<td>1.64</td>
<td>.51</td>
<td>4.38</td>
<td>1.22</td>
</tr>
<tr>
<td>1.68</td>
<td>17.55</td>
<td>1.56</td>
<td>1.95</td>
<td>2.15</td>
<td>2.48</td>
<td>.48</td>
<td>3.28</td>
<td>.81</td>
</tr>
<tr>
<td>1.74</td>
<td>17.97</td>
<td>2.00</td>
<td>2.03</td>
<td>2.00</td>
<td>2.38</td>
<td>.50</td>
<td>3.31</td>
<td>.98</td>
</tr>
<tr>
<td>1.93</td>
<td>14.66</td>
<td>2.88</td>
<td>2.50</td>
<td>2.07</td>
<td>2.32</td>
<td>.48</td>
<td>3.72</td>
<td>1.04</td>
</tr>
<tr>
<td>1.77</td>
<td>17.31</td>
<td>1.36</td>
<td>1.72</td>
<td>2.24</td>
<td>2.25</td>
<td>.52</td>
<td>3.10</td>
<td>.78</td>
</tr>
<tr>
<td>1.94</td>
<td>14.32</td>
<td>2.66</td>
<td>2.53</td>
<td>1.74</td>
<td>2.64</td>
<td>.50</td>
<td>3.48</td>
<td>.93</td>
</tr>
<tr>
<td>1.83</td>
<td>15.05</td>
<td>2.43</td>
<td>1.90</td>
<td>1.46</td>
<td>1.97</td>
<td>.46</td>
<td>3.48</td>
<td>.90</td>
</tr>
<tr>
<td>2.09</td>
<td>15.47</td>
<td>2.42</td>
<td>2.18</td>
<td>.74</td>
<td>2.46</td>
<td>.48</td>
<td>3.16</td>
<td>.86</td>
</tr>
<tr>
<td>1.72</td>
<td>16.85</td>
<td>2.16</td>
<td>2.16</td>
<td>2.84</td>
<td>2.36</td>
<td>.49</td>
<td>3.68</td>
<td>.95</td>
</tr>
<tr>
<td>1.49</td>
<td>17.42</td>
<td>2.12</td>
<td>2.14</td>
<td>3.30</td>
<td>2.04</td>
<td>.48</td>
<td>3.28</td>
<td>1.06</td>
</tr>
<tr>
<td>1.52</td>
<td>18.55</td>
<td>1.87</td>
<td>1.98</td>
<td>2.90</td>
<td>2.16</td>
<td>.48</td>
<td>3.56</td>
<td>.84</td>
</tr>
<tr>
<td>1.64</td>
<td>18.74</td>
<td>2.10</td>
<td>1.89</td>
<td>2.82</td>
<td>2.04</td>
<td>.53</td>
<td>3.56</td>
<td>1.02</td>
</tr>
<tr>
<td>1.40</td>
<td>14.79</td>
<td>2.21</td>
<td>2.07</td>
<td>2.79</td>
<td>2.15</td>
<td>.52</td>
<td>3.49</td>
<td>1.04</td>
</tr>
<tr>
<td>1.78</td>
<td>18.86</td>
<td>2.00</td>
<td>2.08</td>
<td>3.14</td>
<td>2.60</td>
<td>.50</td>
<td>3.30</td>
<td>.80</td>
</tr>
<tr>
<td>1.93</td>
<td>15.62</td>
<td>2.26</td>
<td>2.21</td>
<td>2.81</td>
<td>2.18</td>
<td>.44</td>
<td>4.16</td>
<td>.92</td>
</tr>
<tr>
<td>1.53</td>
<td>18.56</td>
<td>2.14</td>
<td>2.00</td>
<td>3.16</td>
<td>2.22</td>
<td>.51</td>
<td>3.73</td>
<td>1.07</td>
</tr>
</tbody>
</table>

\[ S^2 \text{ or } Sy^2 \]

0.6690 101.4644 6.5921 1.8311 8.8102 1.5818 .0258 3.7248 .3828

Sums of cross-products

<table>
<thead>
<tr>
<th>(x_1)</th>
<th>(x_2)</th>
<th>(x_3)</th>
<th>(x_4)</th>
<th>(x_5)</th>
<th>(x_6)</th>
<th>(y_1)</th>
<th>(y_2)</th>
<th>(y_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.3589</td>
<td>-.0125</td>
<td>-.0244</td>
<td>1.6379</td>
<td>.5057</td>
<td>.2501</td>
<td>-.96105</td>
<td>2.6691</td>
<td></td>
</tr>
<tr>
<td>-.3469</td>
<td>.0352</td>
<td>.7920</td>
<td>.2173</td>
<td>-.5136</td>
<td>12.8511</td>
<td>-2.0617</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.0415</td>
<td>-.14278</td>
<td>-.4753</td>
<td>.5007</td>
<td>2.4054</td>
<td>-.9503</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.0043</td>
<td>.0154</td>
<td>-.0421</td>
<td>.3945</td>
<td>-.4187</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.9120</td>
<td>-.1914</td>
<td>-.93692</td>
<td>3.4020</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.1586</td>
<td>-.32733</td>
<td>1.1663</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(a) Compute the regression of percent nicotine ($Y_3$) on the percentages of total nitrogen ($X_1$) and potassium ($X_3$). Set up the analysis of variance; then determine the estimated error variance and the standard errors of the regression coefficients. Test each regression coefficient for significance. How much is the error variance reduced by using both $X_1$ and $X_3$ rather than $X_1$ alone to estimate $Y_3$? What percentage of the original variability between the $Y_3$'s is accounted for by the regression on $X_1$ and $X_3$?

(b) Analyze the regression of $Y_1$ on $X_2$ and $X_4$ and of $Y_2$ on $X_5$ and $X_6$. 
8.5. General Least Squares for \( \hat{Y} = \bar{Y} + \sum_{i=1}^{r} b_i x_i \).

8.5.1. Introduction:

Now let us suppose that we approximate \( Y \) by means of the general equation

\[
Y = \mu + \sum_{i=1}^{r} \beta_i x_i + \epsilon = \bar{Y} + \sum_{i=1}^{r} b_i x_i + \epsilon
\]

where the number of observations \( n \geq r + 1 \). The usual assumptions which are made are:

(i) The \( \{ x_i \} \) are fixed variates and may be looked upon as population parameters.

Often the \( x_i \)’s are chosen deliberately and the \( Y \)’s are produced or chosen at random.

(ii) For a fixed set of \( x_i \)'s, say \( \{ x_i \} \), the \( Y \)'s associated with this set are NID with mean \( \mu + \sum_{i=1}^{r} \beta_i x_i \) and variance \( \sigma^2 \).

(iii) For any set of \( x_i \)'s, the variance of \( Y \) shall be the same; this is the assumption of homoscedasticity.

The error sum of squares is

\[
SSE = S(y - \sum_{i=1}^{r} b_i x_i)^2.
\]

Again we note that \( \sum \) indicates summation over variables, while \( S \) is for summation over sample values. We are writing \( b_i \) and \( \beta_i \) instead of \( b_{ir} \) and \( \beta_{ir} \).

The general least squares equation by minimizing SSE with respect to \( b_i \) is

\[
\sum_{j=1}^{r} b_j S_{x_j x_i} = S_{x_i y} \quad (i = 1, 2, \ldots, r).
\]

Hence we now have \( r \) equations in the \( r \) unknowns \( \{ b_i \} \).

\[
\begin{align*}
 b_1 S_{x_1} + b_2 S_{x_1 x_2} + \cdots + b_r S_{x_1 x_r} &= S_{x_1 y} = \beta_1 \\
 b_1 S_{x_1 x_2} + b_2 S_{x_2} + \cdots + b_r S_{x_2 x_r} &= S_{x_2 y} = \beta_2 \\
 \vdots \\
 b_1 S_{x_1 x_r} + b_2 S_{x_2 x_r} + \cdots + b_r S_{x_r} &= S_{x_r y} = \beta_r 
\end{align*}
\]
This system of linear equations can be solved by the usual methods of simultaneous equations given in elementary algebra. Methods of solution have been presented by Snedecor (10) and in a special computing manual by Wallace and Snedecor (11). However, we believe that in general it is better to solve for the b's by use of matrix inversion techniques. In this case

\[
A = \begin{bmatrix}
Sx_1^2 & Sx_1 x_2 & \cdots & Sx_1 x_r \\
Sx_1 x_2 & Sx_2^2 & \cdots & Sx_2 x_r \\
\vdots & \vdots & \ddots & \vdots \\
Sx_1 x_r & Sx_2 x_r & \cdots & Sx_r^2
\end{bmatrix}, \quad
B = \begin{bmatrix}
b_1 \\
b_2 \\
\vdots \\
b_r
\end{bmatrix}, \quad
G = \begin{bmatrix}
s_1 \\
s_2 \\
\vdots \\
s_r
\end{bmatrix}
\]

\[
C = A^{-1} = \begin{bmatrix}
c_{11} & c_{12} & \cdots & c_{1r} \\
c_{21} & c_{22} & \cdots & c_{2r} \\
\vdots & \vdots & \ddots & \vdots \\
c_{r1} & c_{r2} & \cdots & c_{rr}
\end{bmatrix}
\]

\[
\sum_{j=1}^{r} c_{ij} Sx_j x_j = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j \end{cases}
\]

The solution for the b's is:

\[
b_1 = \sum_{j=1}^{r} c_{1j} \varepsilon_j = \sum_{j=1}^{r} c_{1j} Sx_j \left( \sum_{k=1}^{r} \beta_k^j x_k + \varepsilon - \bar{\varepsilon} \right)
\]

\[
= c_{11}(\beta_1^1 Sx_1^2 + \beta_2^1 Sx_1 x_2 + \cdots + \beta_r^1 Sx_1 x_r) + \cdots + \\
+ c_{ir}(\beta_1^r Sx_1 x_1 + \beta_2^r Sx_2 x_2 + \cdots + \beta_r^r Sx_r^2) + \sum_{j=1}^{r} c_{ij} Sx_j \varepsilon
\]

\[
= \beta_1^1 (c_{11} Sx_1^2 + \cdots + c_{1r} Sx_1 x_r) + \cdots + \\
+ \beta_1^r (c_{11} Sx_1 x_1 + \cdots + c_{1r} Sx_1 x_r) + \cdots + \\
+ \beta_r^1 (c_{11} Sx_1 x_1 + \cdots + c_{1r} Sx_1 x_r) + \cdots + \\
+ \beta_r^r (c_{11} Sx_1 x_1 + \cdots + c_{1r} Sx_1 x_r) + \sum_{j=1}^{r} c_{ij} Sx_j \varepsilon
\]

\[
= \beta_1^1 + \sum_{j=1}^{r} c_{1j} Sx_j \varepsilon
\]
\[ E(b_1) = \beta_1 \] (indicating \( b_1 \) is an unbiased estimate of \( \beta_1 \))

\[ \sigma^2(b_1) = c_{11} \sigma^2 ; \quad \sigma(b_1b_1') = c_{11} \sigma^2 \]

\[ \sigma^2(b_1 - b_j) = (c_{11} - 2c_{1j} + c_{jj}) \sigma^2 \]

It should be noted that since the \( \varepsilon \) are assumed NID(0, \( \sigma^2 \)), the \( \{ b_i \} \) are multivariate normally distributed with means \( \{ \beta_i \} \), variances \( c_{11} \sigma^2 \) and covariances \( c_{1j} \sigma^2 \).

The error sum of squares is

\[
\text{SSE} = S(y - \sum_{i=1}^{r} b_i x_i)^2
\]

\[
= S y^2 - \left( \sum_{i=1}^{r} b_i S x_i y \right) + \sum_{i=1}^{r} b_i \left( \sum_{k=1}^{r} b_k S x_k x_k - S x_i y \right)
\]

\[
= S y^2 - \left( \sum_{i=1}^{r} b_i S x_i y \right),
\]

because the values in parentheses of the second equation are simply the normal equations, where

\[
\sum_{k=1}^{r} b_k S x_i x_k = S x_i y.
\]

Hence the reduction due to regression is

\[
\text{SSR} = \sum_{i=1}^{r} b_i S x_i y = R^2 S y^2,
\]

where \( R \) is the multiple correlation coefficient.

In a subsequent section, we will prove that

\[
E(\text{SSE}) = (n - r - 1) \sigma^2 ; \quad s^2 = \text{SSE}/(n - r - 1).
\]

\[
E(\text{SSR}) = r \sigma^2 \quad \text{(when all } \beta_i = 0).\]

Hence \( F = \text{SSR}/rs^2 \) can be used to test the null hypothesis that all \( \beta_i = 0 \).

\( F \) has \( (r, n - r - 1) \) degrees of freedom. Also \( \text{SSR} > r \sigma^2 \) when some \( \beta_i \neq 0 \).

\( 1/s^2 \) is denoted as \( s^2_y \) in most discussions of regression analysis.
If it is desired to know if the last \((r - k)\) of the \(r\) independent variates made a significant contribution to SSR, we can obtain the reduction due to the first \(k\) variates by using

\[
\hat{Y} = \bar{Y} + b_1^i x_1 + \cdots + b_k^i x_k.
\]

This reduction will be called SSR\(_k\). Then the added reduction due to the last \((r - k)\) variates is \((SSR - SSR\_k)\). The expected value of \((SSR - SSR\_k)\) is a function of only the last \((r - k)\) \(\beta^i\)'s. Hence we can test the null hypothesis that those \((r - k)\) \(\beta^i\)'s = 0 without saying anything about the first \(k\) variates. The analysis of variance is

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>First (k) variates</td>
<td>(k)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Added reduction by</td>
<td>(r - k)</td>
<td>((SSR - SSR_k)/(r - k)) (s^2 + Q \binom{\beta_{k+1}, \ldots, \beta_r}{r})</td>
<td></td>
</tr>
<tr>
<td>last ((r - k)) variates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>(n - r - 1)</td>
<td>(s^2)</td>
<td>(s^2)</td>
</tr>
</tbody>
</table>

\(Q\) is a function of only \(\{\beta_{k+1}, \beta_{k+2}, \ldots, \beta_r\}\).

Under the null hypothesis \(H_0:\) \(\{\beta_{k+1} = \beta_{k+2} = \cdots = \beta_r = 0\}\), \(Q\) = 0. Hence

\[
\frac{SSR - SSR\_k}{s^2(r - k)} = F
\]

with \((r - k)\) and \((n - r - 1)\) degrees of freedom.

### 8.5.2. Computational Methods:

In general if \(r\) is large, some iterative procedure is needed to solve for the \(c_{ij}\) values. A direct method is all right for \(r\) as small as 4. Snedecor presents a similar method (10) but based on simple correlations instead of the sums of squares and cross products. We define

\[
r_{ij} = \frac{Sx_i x_j}{\sqrt{Sx_i^2, Sx_j^2}}
\]
The $q$ matrix is then replaced by Snedecor's $k$ matrix. R. A. Fisher discusses the use of the $q$ matrix (12). If $r$ is so large that iterative methods are needed, the reader is advised to read an article by Waugh and Dwyer (13). The bibliography for this article contains most of the other articles appropriate to the subject. Some illustrative material has been prepared by Paul Poach of the Institute of Statistics (14).

We will present an example for $r = 4$, using as independent variables all four presented in the table on page 116 for the estimation of Vitamin $B_2$. In order to simplify the computation, we have coded the independent variables as follows:

$$x_1^t = x_1/100, \ x_2^t = x_2/100, \ x_3^t = x_3/10, \ x_4^t = x_4/1000 = x_1^t x_2^t$$

The regression equation is

$$\hat{Y} = \bar{Y} + b_1^t x_1^t + b_2^t x_2^t + b_3^t x_3^t + b_4^t x_4^t.$$

The matrices for these variables are

$$A = \begin{bmatrix}
10.25767 & 0.93798 & 6.87167 & 1.17904 \\
1.1550 & -0.06320 & 1.99828 & 0.166956 \\
6.94667 & 15.94667 & 3.94418 & 0.216976 \\
31.6650 & -36.8374 & 46.7156 & -152.8797
\end{bmatrix}$$

$$G = \begin{bmatrix}
31.6650 \\
-36.8374 \\
46.7156 \\
-152.8797
\end{bmatrix}$$

$$B = \begin{bmatrix}
b_1^t \\
b_2^t \\
b_3^t \\
b_4^t
\end{bmatrix}$$

Two methods of obtaining the inverse matrix will be presented. First the Doolittle method, used by Fisher (12) and Snedecor (10). The computations are presented on page 136.

The procedures followed in these computations were:

1. The original $A$ matrix plus an identity matrix on the right (zeros everywhere except ones on the main diagonal) and a check column (sum of all elements in each row). In all of the computations which follow, the same procedure is followed with the check column. Then if the computing was done correctly, the sum of each row will equal the value in the check column.
### A Matrix

$$
\begin{bmatrix}
    x_1' \\
    x_2' \\
    x_3' \\
    x_4'
\end{bmatrix}
$$

### Identity Matrix

$$
\begin{bmatrix}
    1 & 0 & 0 & 0 \\
    0 & 1 & 0 & 0 \\
    0 & 0 & 1 & 0 \\
    0 & 0 & 0 & 1
\end{bmatrix}
$$

### Check

$$
\begin{bmatrix}
    19,34636 \\
    4,08856 \\
    23,92210 \\
    8,28846
\end{bmatrix}
$$

### II

$$
\begin{bmatrix}
  8.70002 & 0.032213 & 5.82819 & 1 \\
  0.019006 & 0.558230 & -0.031627 & 1 \\
  41.15857 & -0.378543 & 95.51421 & 1 \\
  0.298932 & 0.506640 & 0.042331 & 1
\end{bmatrix}
$$

### Check

$$
\begin{bmatrix}
    0.848148 & 0 & 0 & 0 \\
    0 & 0.500430 & 0 & 0 \\
    0 & 0 & 5.98960 & 0 \\
    0 & 0 & 0 & 0.253538
\end{bmatrix}
$$

### III

$$
\begin{bmatrix}
  8.40109 & -0.474427 & 5.78586 & 1 \\
 -0.279926 & 0.051590 & -0.073958 & 1 \\
  40.85964 & -0.885183 & 95.47188 & 1
\end{bmatrix}
$$

### Check

$$
\begin{bmatrix}
    0.848148 & 0 & 0 & 0 \\
    0 & 0.500430 & 0 & 0 \\
    0 & 0 & 5.989602 & -0.253538
\end{bmatrix}
$$

### IV

$$
\begin{bmatrix}
  1.45200 & -0.081998 & 1 \\
  3.78493 & -0.697558 & 1 \\
  0.42798 & -0.009272 & 1
\end{bmatrix}
$$

### Check

$$
\begin{bmatrix}
    0.146590 & 0 & 0 & 0 \\
    0 & -6.76641 & 0 & 3.428135 \\
    0 & 0 & 0.062737 & -0.002656
\end{bmatrix}
$$

### V

$$
\begin{bmatrix}
  1.02402 & -0.072726 & 1 \\
  3.35695 & -0.688286 & 1
\end{bmatrix}
$$

### Check

$$
\begin{bmatrix}
    0.146590 & 0 & -0.062737 & 0.041164 \\
    0 & -6.76641 & -0.062737 & 3.430791 \\
    0 & 0.091150 & -0.062737 & -0.002656
\end{bmatrix}
$$

### VI

$$
\begin{bmatrix}
  -14.08052 & 1 \\
  -4.87726 & 1
\end{bmatrix}
$$

### Check

$$
\begin{bmatrix}
    -2.01565 & 0 & 0.862649 & 0.566015 \\
    9.83081 & 0.091150 & -4.984454 & 1.06016
\end{bmatrix}
$$

### VII

$$
\begin{bmatrix}
  -9.20326 & 1 \\
  4.67726 & 1
\end{bmatrix}
$$

### Check

$$
\begin{bmatrix}
    -2.01565 & -9.83081 & 0.77150 & 5.55056 \\
    9.83081 & 0.091150 & -4.984454 & -1.47276
\end{bmatrix}
$$

### VIII

$$
\begin{bmatrix}
  (1)c_{11} & c_{12} & c_{13} & c_{14} \\
  c_{21} & (1)c_{22} & c_{23} & c_{24} \\
  c_{31} & c_{32} & (1)c_{33} & c_{34} \\
  c_{41} & c_{42} & c_{43} & (1)c_{44}
\end{bmatrix}
$$

### Check

$$
\begin{bmatrix}
    0.219015 & 1.06819 & -0.83829 & 0.693108 \\
    1.06819 & 15.04066 & -3.17707 & -7.92605 \\
    -0.83829 & -3.17707 & 0.95668 & 1.81973 \\
    -0.603108 & -7.92605 & 1.81973 & 4.44178
\end{bmatrix}
$$

### XII

$$
\begin{bmatrix}
  g_1 \\
  b_1
\end{bmatrix}
$$

### Check

$$
\begin{bmatrix}
    31.6650 & -16.8374 & 46.7156 & -152.8797 \\
    2.4631 & -75.3773 & 1.5836 & -1.3769
\end{bmatrix}
$$

### XIV

$$
\begin{bmatrix}
  s(b_{11}') = \sqrt{c_{11}s^2} \\
  b_{11}
\end{bmatrix}
$$

### XV

$$
\begin{bmatrix}
  t \\
  \leq 1 & 1.93 & \leq 1 & \leq 1
\end{bmatrix}
$$

### XVI

\( SSR = 6908.04 \); \( SSE = 2242.49 \); \( s^2 = 101.93 \)
II. Divide each row by the element in the $x_i$ column of the $A$ matrix. Be sure to carry at least 5 significant digits, and preferably 6, in all quotients. Remember that the important point is the number of significant digits and not the number of decimal places.

III. Subtract the last line from all the others, dropping the $x_i$ column in the left-hand matrix.

IV. Divide by the elements of the right column in this new left-hand matrix.

V. Again subtract the last line from the others and now drop the $x_j$ column.

VI, VII. Same as above.

VIII. The first line gives the $c_{1j}$ values, which are computed by dividing the elements in VII by the left-hand number, $-9.20326$.

IX. Substitute in turn the $c_{1j}$ values in either line of the left-hand matrix of VI as follows to determine $c_{2j}$. For example

$$c_{21} = \begin{cases} 
-2.01565 - (-14.08052)(0.219015) \\
\text{or} \\
0 - (-4.87726)(0.219015)
\end{cases} = 1.06819.$$

Note that $c_{21}$ should equal $c_{12}$, except for rounding errors.

$$c_{22} = \begin{cases} 
0 - (-14.08052)(1.06819) \\
9.83081 - (-4.87726)(1.06819)
\end{cases} = 15.04066$$

X. Next substitute the values of $c_{1j}$ and $c_{2j}$ in any of the equations in IV and solve for $c_{3j}$ as in IX for $c_{2j}$. For example

$$c_{31} = \begin{cases} 
.146590 - (1.45200)(0.219015) - (-.081998)(1.06819) = -.083830 \\
0 - (3.78493)(0.219015) - (-.697558)(1.06819) = -.083832 \\
0 - (0.42798)(0.219015) - (-.009272)(1.06819) = -.083830
\end{cases}$$

We note that two of these equations give the same result to 5 significant digits (6 decimal places) while the middle one deviates very slightly from these two. We have used the value given in VIII for $c_{13}$. Rounding errors are quite a problem in matrix inversion calculations. Hence it is advisable to carry several unnecessary digits at first in order to be able to drop
digits as the computing proceeds and end up with as many as you thought necessary.

XI. Next substitute the values of $c_{1j}$, $c_{2j}$, and $c_{3j}$ in any of the equations in II and solve for $c_{4j}$. For example

$$c_{41} = \begin{cases} 0.848148 - 8.70002c_{11} - 0.032213c_{21} - 5.82819c_{31} = -0.603125 \\ 0 - 0.019006c_{11} - 0.558230c_{21} + 0.031627c_{31} = -0.603110 \\ 0 - 41.15857c_{11} + 0.378543c_{21} - 95.51421c_{31} = -0.603128 \\ 0 - 0.298932c_{11} - 0.506640c_{21} - 0.042331c_{31} = -0.603111 \end{cases}$$

In this case two equations are cut of line and two almost agree with $c_{4j}$.

In general it is a good practice to omit equations like number 1 and number 3, because several of the elements are so large. These large coefficients (such as 41 and 95) are subject to a much greater rounding error, because extra digits are needed to give 6 decimal accuracy. That is, we are trying to calculate $c_{41}$ to 6 decimal places; hence, the individual multipliers should be accurate to 6 decimal places. But for numbers like 41 and 95 to have 6 decimal places, 8 significant digits are required, and we cannot obtain 8 significant digits in our computing unless we use several more in the original A matrix.

The C-matrix has now been completed but the computer should have some more checking besides the check column. The final check is to find if the product of the A and C matrices is the identity matrix. That is

$$AC = D = I$$

where I consists of only ones in the main diagonal. To compute an element in the $i^{th}$ row and $j^{th}$ column of $AC$, we find the sums of the products of corresponding elements in the $i^{th}$ row of $A$ and the $j^{th}$ column of $C$. Let $a_{ik}$ be the elements in $A$ and $c_{kj}$ those in $C$. Then the $(i, j)$ element in $D$ is

$$d_{ij} = \sum_{k=1}^{4} a_{ik} c_{kj}$$
We generally compute the diagonal elements \( d_{11} \) to see if they are nearly one (to the desired degree of accuracy). In the Vitamin B₂ example

\[
\begin{align*}
d_{11} &= (10.25767)(0.219015) + \cdots + (1.17904)(-0.603108) = 1.0000198 \\
d_{22} &= (0.03798)(1.06819) + \cdots + (1.99828)(-7.92605) = 1.0000380 \\
d_{33} &= 1.0000014 \\
d_{44} &= 1.0000057
\end{align*}
\]

The last two diagonal elements are within the desired five decimal place accuracy but \( d_{11} \) and \( d_{22} \) only to 4 places. Hence if we use the present C-matrix, we cannot hope for more than 4-place accuracy in the b's and possibly only 3-place accuracy.

XII. The sums of the cross-products of the \( x_i \) with \( y \) are given here (\( g_i = Sx_iy \)).

XIII. \( b_i^1 = \sum_{j=1}^{4} c_{ij}g_j \). For example,

\[
b_1^1 = (0.219015)(31.6650) + \cdots + (-0.603108)(-152.8797) = 2.4631
\]

At this stage it is advisable to either recompute the \( b_i^1 \) or substitute these values in the original normal equations. For example,

\[
10.25767b_1^1 + \cdots + 1.17904b_4^1 = 31.6650
\]

Substituting the above values of the \( b_i^1 \), this sum is 31.6644, indicating slight inaccuracies. The exact values are given by use of \( C'' \) below.

XVI. \( SSR = \sum b_i^1 g_i = (2.4631)(31.6650) + \cdots + (-1.3769)(-152.8797) = 6908.04 \)

\( SSE = SY^2 - SSR = 9150.53 - 6908.04 = 2242.49 \)

\( s^2 = SSE/22 = 101.93 \)

XIV. The standard error of \( b_i^1 = \sqrt{c_{ii}s^2} \)

XV. \( t = b_i^1/s(b_i^1) \)

If it is desired to improve the results to more significant figures, we can use an iterative device advanced by Hotelling (15). The procedure is as follows:
(1) Compute the matrix \((2 - AC)\), which is the same as \(AC\), except the diagonal elements are subtracted from 2 and the signs of all other elements are reversed.

\[
(2 - AC) = \begin{bmatrix}
0.9999802 & -0.0000771 & 0.000044 & 0.000195 \\
-0.000035 & 0.9999620 & -0.000008 & -0.000046 \\
-0.000033 & -0.000092 & 0.999986 & -0.000124 \\
0.000079 & -0.000778 & -0.000013 & 0.999943
\end{bmatrix}
\]

(2) Now compute \(C(2 - AC) = C''\), the new C-matrix

\[
C'' = \begin{bmatrix}
0.219012 & 1.068180 & -0.83828 & -6.03104 \\
1.068180 & 15.040626 & -3.17704 & -7.926049 \\
-0.83828 & -3.17704 & 0.95668 & 1.181971 \\
-6.03104 & -7.926049 & 1.81971 & 4.441777
\end{bmatrix}
\]

(3) \(b''_4 = \begin{bmatrix}
2.46332 \\
-75.3747 \\
1.58369 \\
-1.37645
\end{bmatrix}\)

(4) Substituting in the normal equations

\[
\begin{bmatrix}
31.6649 & -86.8375 & 46.7156 & -152.8800
\end{bmatrix}
\]

as compared to the exact values

\[
\begin{bmatrix}
31.6650 & -86.8374 & 46.7156 & -152.8797
\end{bmatrix}
\]

(5) If we use these values of \(b''_4\), we find that \(SSR = \sum b''_4 x_4 = 6907.76\) as compared to 6908.04. Hence there is no real difference in the results by using the \(b'_4\).

Now we shall present a short-cut method of inverting a symmetric matrix \((a_{ij} = a_{ji})\), the abbreviated Doolittle method, described by Dwyer (16).

\[
\begin{array}{cccccc}
& x_1 & x_2 & x_3 & y & \text{Check} \\
\hline
I & \begin{bmatrix} 10.25767 \\
1.1150 \\
15.94667 \\
3.94418 \\
3.08469 \end{bmatrix} & \begin{bmatrix} 0.037980 \\
-0.063200 \\
0.166956 \\
-152.8797 & -145.5912
\end{bmatrix} & \begin{bmatrix} 1.17904 \\
0.099667 \\
0.166956 \\
6.7156 \\
-145.5912 \end{bmatrix} & \begin{bmatrix} 31.6650 \\
-86.8374 & 46.7156 & -152.8797 & 69.6377 & 69.6377
\end{bmatrix} & \begin{bmatrix} 50.0114 \\
83.7488 \\
69.6377 \\
-145.5912 & \end{bmatrix}
\end{array}
\]

\[
\begin{array}{cccccc}
II & A_1 & 10.25767 & 0.037980 & 1.17904 & 31.6650 & 50.0114 \\
B_1 & 1 & 0.00370260 & 0.669906 & 0.114942 & 3.08469 & 4.87551
\end{array}
\]

\[
\begin{array}{cccccc}
III & A_2 & 1.09536 & -0.088643 & 1.99391 & -86.9546 & -83.9340 \\
B_2 & 1 & -0.079475 & 1.78768 & -77.9610 & -75.2528
\end{array}
\]

\[
\begin{array}{cccccc}
IV & A_3 & 11.03625 & -0.464424 & 18.5923 & 29.4641 \\
B_3 & 1 & -0.040968 & 1.64007 & 2.59910
\end{array}
\]

\[
\begin{array}{cccccc}
V & A_4 & 0.22516 & -0.3104 & -0.0852 \\
B_4 & 1 & -1.3786 & -0.3784
\end{array}
\]

\[
\begin{array}{cccccc}
& \text{Backward Solution (for } a_{ij} \text{)} & \\
\hline
a_{11} & 0.219003 & 1.06806 & -0.0838254 & -0.603037 \\
a_{21} & 15.03894 & -0.317667 & -7.92514 \\
a_{31} & 0.956668 & 0.181951 & 4.44129
\end{array}
\]
The procedures followed in the abbreviated Doolittle method computations were:

I. The original $A$ matrix with only the upper right corner is reproduced plus the $G$ column ($S_{4j}^I$) and a check column. For the check column, we assume the entire $A$ matrix is present. Since the $A$ matrix is symmetrical, the lower corner is an image of the upper corner. We shall call these elements $a_{ij}$.

II. $a_{1j} = a_{1j}$, the first row of $I$ is reproduced

$$B_{1j} = A_{1j} / A_{11} = A_{1j} / 10.25767.$$  

III. $a_{2j} = a_{2j} - \left[ A_{12}B_{1j} \ or \ A_{1j}B_{12} \right]$, where $a_{2j}$ is the element in the second row of $I$ and $A_{12}B_{1j} = A_{1j}B_{12}$ except for rounding errors. As we noted in commenting on rounding errors for the Doolittle method, it is advisable to choose from these two ($A_{12}B_{1j}$ or $A_{1j}B_{12}$) the one for which the two members are more nearly equal.

$$A_{22} = 1.11550 - (.037980)(.0037026) = 1.11536$$

$$A_{23} = -.063200 - \left\{ {\begin{array}{l} (.037980)(.669906) = -0.088643 \\
(.0037026)(6.87167) = -0.088643 \end{array}} \right.$$  

$$B_{2j} = A_{2j} / A_{22}$$

IV. $a_{3j} = a_{3j} - (A_{13}B_{1j} + A_{23}B_{2j}) = a_{3j} - (A_{1j}B_{13} + A_{2j}B_{23})$

$$A_{33} = 15.94667 - \left[ (6.87167)(.669906) + (-.088643)(-.079475) \right] = 11.33625$$

$$A_{34} = 0.166956 - \left\{ {\begin{array}{l} (6.87167)(1.14942) + (-.088643)(1.78768) = -0.464422 \\
(1.17904)(.669906) + (-.079475)(1.99391) = -0.464424 \end{array}} \right.$$

$$B_{3j} = A_{3j} / A_{33}.$$  

V. $a_{4j} = a_{4j} - (A_{14}B_{1j} + A_{24}B_{2j} + A_{34}B_{3j}) = a_{4j} - (A_{1j}B_{14} + A_{2j}B_{24} + A_{3j}B_{34}).$

$$A_{44} = 3.94418 - \left[ (1.17904)(1.14942) + (1.99391)(1.78768) + (.464424)(.040968) \right] = 0.22516$$

$$B_{4j} = A_{4j} / A_{44}.$$  

This completes the forward solution. If the experimenter only desires to compute the $b^I_j$ and the over-all reduction in sum of squares due to regression, SSR,
without the individual \( s^2(b_{ij}) \) and \( s(b_{ij}^2) \), he can make those computations without determining the inverse matrix, as follows:

\[
b_4' = B_{4y} = -1.3786
\]

\[
b_3' = B_{3y} = B_{3y}^i = 1.64007 + (.040968)(-1.3786) = 1.5836
\]

\[
b_2' = B_{2y} = B_{23}b_3' - B_{24}b_4' = -77.9610 - (-.079475)(1.5836) - (1.78768)(-1.3786)
\]

\[= -75.3706
\]

\[
b_1' = B_{1y} = B_{12}b_2' - B_{13}b_3' - B_{14}b_4' = 3.08696 - (.0037026)(-75.3706) - (.669906)(1.5836)
\]

\[= (.114942)(-1.3786) = 2.4636
\]

\[
SSR = \sum b_1'y_i = (2.4636)(31.6650) + \cdots + (-1.3786)(-152.8797) = 6907.74
\]

We note that \( B_{1y} = b_{11}' \) and that \( A_{1y} = Sx_1'y_i, A_{2y} = S(x_2' \text{ adjusted for } x_1')y_i, A_{3y} = S(x_3' \text{ adjusted for } x_1' \text{ and } x_2')y_i, \) and \( A_{4y} = S(x_4' \text{ adjusted for } x_1', x_2', \text{ and } x_3')y_i \).

Following our results for two independent variables, we can consider the regression equation as

\[
y = b_{11}'x_1' + b_{22}'(x_2' \text{ adjusted for } x_1') + b_{33}'(x_3' \text{ adjusted for } x_1' \text{ and } x_2') + \\
b_{44}'(x_4' \text{ adjusted for } x_1', x_2', \text{ and } x_3') + o_4
\]

Hence we can also write

\[
SSR_4 = \sum_{i=1}^{4} A_{1y}b_{11}' = \sum_{i=1}^{4} A_{1y}B_{1y} = (31.6650)(3.08696) + \cdots + (-0.3104)(-1.3786) = 6907.74
\]

Despite the obvious errors of rounding in computing the \( b_i' \) by this short-cut procedure, SSR is only .02 less than the exact value computed after the use of the Hotelling iterative procedure on the Doolittle solutions.

If the computer needs the inverse (0) matrix, the computations proceed as given for the so-called backward solution.

I. First compute \( a_{ij} \) values.

\[
a_{44} = 1/a_{44} = 1/0.44129 = 4.44129
\]

\[
a_{34} = -a_{44}B_{34} = -(4.44129)(-0.049068) = 0.181951
\]

\[
a_{24} = -a_{34}B_{23} - a_{44}B_{24} = (.181951)(.079475) - (4.44129)(1.78768) = -7.92514
\]

\[
a_{14} = -a_{24}B_{12} - a_{34}B_{13} - a_{44}B_{14} = -6.03037
\]
II. Next the $c_{13}$ values

\[ c_{43} = c_{34} = 0.181951 \]
\[ c_{33} = 1/A_{33} - c_{34}B_{34} = 0.0882136 - (0.181951)(0.040968) = 0.0956668 \]
\[ c_{23} = -c_{33}B_{23} - c_{34}B_{24} = (0.0956668)(0.079475) - (0.181951)(1.78768) = -0.317667 \]
\[ c_{13} = -c_{23}B_{12} - c_{33}B_{13} - c_{34}B_{14} = -0.0838254. \]

Again check by use of

\[ a_{13}c_{13} + \cdots + a_{43}c_{43} = 1.0000008. \]

III. $c_{12}$ values

\[ c_{42} = c_{24} = -7.92514 \]
\[ c_{32} = c_{23} = -0.317667 \]
\[ c_{22} = 1/A_{22} - c_{23}B_{23} - c_{24}B_{24} \]
\[ = 0.8965715 - (-0.317667)(-0.079475) + (7.92514)(1.78768) = 15.03894 \]
\[ c_{12} = -c_{22}B_{12} - c_{23}B_{13} - c_{24}B_{14} = 1.06806 \]

and

\[ c_{12}a_{12} + \cdots + c_{42}a_{42} = 0.99993. \]

IV. $c_{11} = 1/A_{11} - c_{12}B_{12} - c_{13}B_{13} - c_{14}B_{14} \]

\[ = 0.974880 - (1.06806)(.0037026) + (.0838254)(.669906) + \]
\[ + (.603037)(.114942) = 0.219003 \]

and

\[ c_{11}a_{11} + \cdots + c_{41}a_{41} = 1.0000002 \]

V. The solutions for the $b_{i}^{1}$, SSR, and $s^{2}(b_{i}^{1})$ proceed as with the Doolittle method

\[ b_{i}^{1} = \begin{bmatrix} 2.46334 & -75.3693 & 1.58356 & -1.37975 \end{bmatrix} \]

\[ SSR = 6907.79 \]

Again we note that some of the $b_{i}^{1}$ are not too accurate but that SSR is off only slightly in the last decimal place. The Hotelling iterative procedure can be used to improve this C-matrix also. Rounding errors seem to be of more importance with the abbreviated method; hence, it is especially advisable to carry extra places
at first in order to secure the desired accuracy at the end without having to use
the Hotelling iterative device to improve the accuracy.

Referring back to the t-values for the regression coefficients (using the
Doolittle method solution), we note that none of these regression coefficients is
significant, although $b_2$ is almost significant. This is a very perplexing con-
cclusion, especially since the regression coefficient for $X_2$ was highly significant
when only $X_1$ and $X_2$ were used as predictors and when $X_2$ alone was used as a
predictor. It is even more perplexing when we consider the analysis of variance.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression (4 variates)</td>
<td>4</td>
<td>6907.76</td>
<td>1726.94**</td>
</tr>
<tr>
<td>Error</td>
<td>22</td>
<td>2242.77</td>
<td>101.94</td>
</tr>
<tr>
<td>Regression ($X_2$ only)</td>
<td>1</td>
<td>6759.98</td>
<td>6759.98</td>
</tr>
<tr>
<td>Added Reduction due to $X_1$, $X_3$ and $X_4$</td>
<td>3</td>
<td>147.78</td>
<td>49.26</td>
</tr>
</tbody>
</table>

From this analysis of variance, we conclude that the over-all reduction due to
the use of the regression equation is highly significant ($F = 1726.94/101.94 = 16.94$)
and that the added reduction due to the three variates other than $X_2$ is decidedly
non-significant. In fact even if the added reduction were attributable to one
variate (147.78 with 1 degree of freedom), the reduction would still be only
slightly greater than $s^2$. Hence we are led to conclude that $X_2$ is a very important
predictor and that the other predictors add nothing to the reliability of the
estimate of Vitamin B2 content.

Why then is $b_2^i$ not highly significant? The answer lies in the peculiar nature
of $X_4^i$, which is the product of $X_1^i$ and $X_2^i$, causing $X_4^i$ and $X_2^i$ to be highly
correlated. Hence $b_2^i$ and $b_4^i$ are also highly correlated, so that the actual
influence of $X_2^i$ on $Y$ is split into a part contributed by $b_2^i$ and another part
contributed by $b_4^i$. It is impossible to interpret $b_2^i$ as the change in $Y$ when $X_2^i$
varies while holding the other $X_1^i$ constant, because $X_4^i$ will vary when $X_2^i$ varies.
The change in $\hat{Y}$ when $X_2^i$ varies is given by

$$ z = \frac{\partial \hat{Y}}{\partial X_2^i} = b_2^i + b_4^ix_1^i. $$
The average value of this change is
\[ \bar{z} = b_2 \bar{x}_2 + b_4 \bar{x}_4 = -75.3747 - 1.3765(1.8578) = -77.9320. \]

This average change is almost the same as \( b_{21} \), the average change in \( \hat{Y} \) for a unit change in \( x_2 \) neglecting all other variables. The estimated variance of this average change, \( \bar{z} \), is
\[
s^2(\bar{z}) = s^2(b_2) + (\bar{x}_2)^2 s^2(b_4) + 2\bar{x}_2 s(b_2 b_4)
= \left[ 15.04063 + 4.44178(1.8578)^2 - 2(1.8578)(7.92605) \right] (101.94)
= (0.92105)(101.94) = 93.893
\]
and the standard error is 9.6899, so that
\[ t = 77.93/9.69 = 8.04, \]
a highly significant value. We have used (2) and (3), page 142, in this computing.

This example was selected because it illustrates some of the difficulties of interpreting regression analyses when some of the independent variates are closely related to one another. If two variates are highly correlated, it seems highly unrealistic to assume that one can be held constant while the other varies. A multiple regression coefficient can only be interpreted as the average change in \( \hat{Y} \) for a unit change in \( x_1 \) when the other \( x \)'s are not changed. Hence to interpret two highly correlated regression coefficients, we would like to know the relationship between the \( x \)'s in order to study the real change in \( \hat{Y} \) when one of the \( x \)'s changes. This is not to say that the use of a variable like \( x_4 = x_1 x_2 \) is undesirable. On the contrary, it is often quite desirable to be able to say how \( \bar{z} \) differs for various values of \( x_1 \). For example, if \( x_1 \) were temperature and \( x_2 \) were rainfall, then it would be highly desirable to know how the effect of one inch of rainfall varied for different temperatures, \( x_1 \). We know that for most crops high rainfall may be detrimental at low temperatures but quite valuable at high temperatures. Hence a knowledge of the regression of yield on temperature and rainfall alone would be rather useless unless the cross-product term were also included. See reference (17) for an example of such a study.
The estimated variance of the average value of $Y$ for a fixed set of $X$'s, \( \{x_i\} \) can be computed directly from the \( \{c_{ij}\} \) and the value of \( s^2 \). Given
\[
\hat{Y} = \bar{Y} + \sum_{i=1}^{r} b_i x_i
\]
\[
s^2(\hat{Y}) = s^2 \left[ \frac{1}{n} + \sum_{i=1}^{r} c_{ii} x_i^2 + 2 \sum_{i < j} c_{ij} x_i x_j \right] = s^2 \left[ \frac{1}{n} + \sum_{i,j=1}^{r} c_{ij} x_i x_j \right],
\]
where \( x_i = X_i - \bar{X} \). Using the values of the \( c_{ij} \), found by the use of the Hotelling iterative procedure (page 142), we have for our Vitamin B2 example
\[
s^2(\hat{Y}) = 101.94 \left[ \frac{1}{27} + x_1(0.219012 x_1 + 1.068180 x_2 - 0.033828 x_3 - 0.03104 x_4) + x_2(1.068180 x_1 + 15.040626 x_2 - 0.317704 x_3 - 7.926049 x_4) + x_3(-0.033828 x_1 - 0.317704 x_2 + 0.095668 x_3 + 0.181971 x_4) + x_4(-0.03104 x_1 - 7.926049 x_2 + 0.181971 x_3 + 4.441777 x_4) \right].
\]
Similarly the estimated variance of a single predicted value of $Y$ is given by
\[
\hat{s}^2(\hat{Y}) + s^2.
\]
Confidence limits can be assigned to the various estimates as follows:
\[
b_1 - t_{\alpha} s(b_1) \leq \beta_1 \leq b_1 + t_{\alpha} s(b_1)
\]
\[
\hat{Y} - t_{\alpha} \hat{s}(\hat{Y}) \leq E \left[ Y \mid \{x_i\} \right] \leq \hat{Y} + t_{\alpha} \hat{s}(\hat{Y})
\]
\[
\hat{Y} - t_{\alpha} \sqrt{s^2(\hat{Y}) + s^2(Y)} \leq Y \mid \{x_i\} \leq \hat{Y} + t_{\alpha} \sqrt{s^2(\hat{Y}) + s^2},
\]
where
\[
P(\mid t > t_{\alpha}) = \alpha.
\]
One final word about this regression analysis. Suppose it were desired to know if the last two variates, \( X_3 \) and \( X_4 \), improved the prediction over the use of \( X_1 \) and \( X_2 \). Then we would set up the following analysis of variance:

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression on ( X_1 ) and ( X_2 )</td>
<td>2</td>
<td>6,876.85</td>
<td>_</td>
</tr>
<tr>
<td>Added reduction due to ( X_3 ) and ( X_4 )</td>
<td>2</td>
<td>30.91</td>
<td>15.46</td>
</tr>
<tr>
<td>Error</td>
<td>22</td>
<td>2,242.77</td>
<td>101.94</td>
</tr>
</tbody>
</table>

The added reduction is definitely not significant.

Exercise 8.5.1. What changes will be made in the regression analysis if the model is \( E(Y) = \sum_{i=1}^{k} \beta_i X_i \)?

Exercise 8.5.2. Each member of the class might be asked to select some data in his own field of application with one dependent and at least three independent variates and to carry out the calculations leading to the estimates of the regression coefficients and their standard errors. Investigate the usefulness of the various independent variates and indicate if any should be discarded.

Exercise 8.5.3. A cost study was made of 89 dairy farms in Central North Carolina in 1941 (18). The dependent variable was the amount of milk sold \( (Y) \) with the following independent variables - amount of concentrates \( (X_1) \), amount of silage \( (X_2) \), pasture cost \( (X_3) \), and amount of roughage \( (X_4) \). The means and sums of squares and cross-products adjusted for the means are given below, with the amounts in tons per cow and the pasture cost in ten dollars per cow.

<table>
<thead>
<tr>
<th>( X_1 )</th>
<th>( X_2 )</th>
<th>( X_3 )</th>
<th>( X_4 )</th>
<th>( Y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
<td>50.5154</td>
<td>-66.1617</td>
<td>-4.84289</td>
<td>0.937732</td>
</tr>
<tr>
<td>( x_2 )</td>
<td>967.1077</td>
<td>13.5895</td>
<td>32.4425</td>
<td>39.0556</td>
</tr>
<tr>
<td>( x_3 )</td>
<td>12.5457</td>
<td>-12.5195</td>
<td></td>
<td>7.02815</td>
</tr>
<tr>
<td>( x_4 )</td>
<td></td>
<td>192.3053</td>
<td></td>
<td>9.99432</td>
</tr>
</tbody>
</table>

\( y \) | 113.5872|

| Means   | 2.94310  | 3.90647  | 1.16426  | 3.60326  | 5.73994 |
(a) Compute the regression coefficients and their standard errors. Use the abbreviated Doolittle method, carrying the \( Y \) column along, but complete the backward solution also.

(b) Show that all regression coefficients except \( b_4 \) are significant at the 1% probability level.

(c) Use the \( Y \) column in the forward solution to determine the error variance when \( X_4 \) is removed from the prediction equation.

(d) In 1941 the cost per ton was $18.00 for concentrates, $2.70 for silage, and $8.50 for roughage and milk sold for 3.2 cents per pound. Which, if any, of the feeds would be adjudged profitable?

(e) Estimate the profit or loss and its standard error if 3200 pounds of concentrates, 4,000 pounds of silage, 3,000 pounds of roughage, and $15.00 worth of pasture were used per cow.

Exercise 8.5.4. Use the data in exercise 8.4.3 to analyze:

(a) The effect on cigarette burn of percent potassium and percent phosphorus and chlorine, and percent calcium and magnesium.

(b) The effect on percent sugar of percent nitrogen and chlorine as well as percent calcium and magnesium.

Exercise 8.5.5. J. T. Wakeley of the Institute of Statistics, North Carolina State College, has compiled some soil weather data and data on the vitamin content of turnip greens contributed by the Georgia Agricultural Experiment Station for the Southern Cooperative Group (1949). The dependent variables are mgs. of ascorbic acid \( [Y_1] \) and mgs. of riboflavin \( [Y_2] \), each per 100 mgs. of dry weight.

The independent variables are soil moisture tension (atmospheres \( \div 10 \)) \( [X_1] \) and mean temperature (degrees F \( \div 100 \)) \( [X_2] \) each at 8" depth, total radiation in gram calories/(cm)^2/min \( \div 1000 \) \( [X_3] \) and evaporation in cms. \( [X_4] \) each for the previous 48 hours, and number of days since planting \( \div 100 \) \( [X_5] \).
The means and sums of squares and cross-products are

<table>
<thead>
<tr>
<th></th>
<th>x₁</th>
<th>x₂</th>
<th>x₃</th>
<th>x₄</th>
<th>x₅</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>x₁</td>
<td>6.6510</td>
<td>2.6722</td>
<td>4.4957</td>
<td>2.6338</td>
<td>0.14360</td>
<td>0.5475</td>
</tr>
<tr>
<td>x₂</td>
<td>4.8750</td>
<td>8.3609</td>
<td>4.5397</td>
<td>-3.2533</td>
<td>2.7156</td>
<td></td>
</tr>
<tr>
<td>x₃</td>
<td>19.8174</td>
<td>9.5315</td>
<td>-6.6767</td>
<td>1.4353</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x₄</td>
<td>5.2394</td>
<td>-3.1452</td>
<td>0.5810</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x₅</td>
<td>3.7799</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ y₁ = -0.60160 \quad -1.5771 \quad -1.6953 \quad -0.81984 \quad 0.94173 \quad 2.8497 \]

\[ y₂ = -3.2590 \quad 7.1768 \quad 10.9862 \quad 5.9285 \quad -7.1415 \quad 6.1834 \]

\[ S_{y₁}² = 1.5025 \quad S_{y₂}² = 28.6900 \quad n = 32 \]

Since the problem of computing a regression with 5 independent variables is quite long, the instructor may want to select particular variables for class work.

**Exercise 8.5.6.** Given the regression model

\[ Y = \mu + \beta_1 x_1 + \cdots + \beta_r x_r + \varepsilon = \hat{Y} + c. \]

Show that

(a) \( \varepsilon \) is uncorrelated with any of the \( x_i \), that is \( S_{x1}c = 0 \) (for any \( i \)).

(b) \( S_{y0} = 0 \)

(c) \( Sc² = SSE = S_{y}² - S_{x}² \)

**Exercise 8.5.7.** Use the orthogonal transformations given in the next section 8.5.3. for \( r = 5 \). [Note that \( \sigma^2 (b_i^1) = \sigma^2 \) and \( \sigma^2 (b_i^1 b_j^1) = 0 \).] Show that

(a) \( b_5 = w_{5,5}b_5^1 \) and \( b_4 = w_{4,4}b_4^1 + w_{5,4}b_5^1 \).

(b) The added reduction in residual sum of squares due to the last two variates (\( X_4 \) and \( X_5 \)) over the reduction due to the first three variates (\( X_1, X_2, \) and \( X_3 \)) alone is given by

\[
\frac{w_{5,5}²b_5² - 2w_{5,4}w_{5,5}b_4b_5 + (w_{4,4}² + w_{5,4}²) b_5²}{w_{4,4}w_{5,5}} = \frac{(c_{5,5}b_5² - 2c_{4,5}b_4b_5 + c_{4,4}b_4²)}{c_{4,4}c_{5,5} - c_{4,5}²}
\]
3.5.3. Theory of Tests of Significance

For references to the theory of tests of significance with regression problems see Bartlett (19), Yates (20) and R. A. Fisher (21).

The equation $Y = \mu + \sum \beta_i x_i + \varepsilon = \bar{Y} + \sum b_i x_i + e$ can be replaced by a new equation

$$Y = \mu + \sum \beta_i' z_i + \varepsilon = \bar{Y} + \sum b_i' z_i + e$$

where $\{z_i\}$ are functions of $\{x_i\}$ so constructed that $\{z_i\}$ are completely orthogonal variates. As before, the $\varepsilon$ are NID(0, $\sigma^2$). Hence the $Y$'s are NID($\mu + \sum \beta_i' z_i$, $\sigma^2$). We might write

$$z_1 = w_{11} x_1$$
$$z_2 = w_{21} x_1 + w_{22} x_2$$
$$\vdots$$
$$z_r = w_{r1} x_1 + w_{r2} x_2 + \cdots + w_{rr} x_r$$

where

$$Sx_i^2 = 1 \text{ and } S(z_i z_j) = 0 \quad (i \neq j)$$

Hence

$$Y = \mu + \sum \beta_i' x_i + \varepsilon = \mu + \sum \beta_i' z_i + \varepsilon$$

where $\{x_i\}$ are solved backwards in term of $\{z_i\}$. The least squares estimate of $\beta_i'$ is $b_i' = S_{z_1 y} = S_{z_1} y$ because of the orthogonal relationships.

$$SSE = S(y - \sum b_i' z_i)^2 = Sy^2 - \sum_{i=1}^{r} (b_i')^2$$

and

$$\sum (b_i')^2$$

is the reduction in sum of squares due to the regression.

---

This section can be omitted if the instructor is not interested in a more theoretical presentation than in the previous sections.
Before continuing, we should show that our solution to the least squares equation is unique. We know that
\[ \hat{Y} = \bar{Y} + \sum_{i=1}^{r} b_i^* z_i = \bar{Y} + \sum_{i=1}^{r} b_i^* x_i \]
where
\[ b_i^* = \sum_{j=1}^{r} w_{ji} b_j^i \cdot \]
For example,
\[ b_1^* = w_{11} b_1^1 + w_{21} b_1^2 + \cdots + w_{rl} b_1^r \cdot \]
But \( b_i^* \equiv b_i \), because both were derived by minimizing \( \sum e^2 \). There can be but one minimum because the \( \sum e^2 \) equation is quadratic in the variables. Hence
\[ \sum (b_i^1)^2 = \sum b_i S_{xy} \]
We shall use the orthogonalized regression coefficients \( \{ b_i \} \) in the theory which follows, always remembering that any \( x_j \) is a function of only \( \{ z_1, z_2, \cdots, z_j \} \).

Because of the relationships
\[ S_{x_i z_j} = \begin{cases} 0 & i \neq j \\ \ell & i = j \end{cases} \]
\[ S_{x_i}(Y - \mu - \sum_{j=1}^{r} \beta_j^i z_j) = S_{x_i} \varepsilon = S_{x_i} Y - \sum_{j=1}^{r} \beta_j^i S_{x_i z_j} = b_i^1 - \beta_i^1 \cdot \]
Hence
\[ (b_i^1 - \beta_i^1) = S_{x_i} \varepsilon \]
a completely orthogonal form in the \( \varepsilon \), which are \( \text{NID}(0, \sigma^2) \). From our theory on orthogonal forms, we know that
\[ \text{E}(\varepsilon) = 0 \text{ or } \text{E}(b_i^1) = \beta_i^1 \]
\[ \sigma^2 = \text{E}(b_i^1 - \beta_i^1)^2 = S_{x_i} \sigma^2 \]
\[ \frac{\text{E}}{\sum} \left[ (b_i^1 - \beta_i^1)(b_j^1 - \beta_j^1) \right] = 0 \quad (i \neq j) \text{.} \]
Since the $e_i$ are NID so are $(b_1^j - \beta_1^j)$. Also $t = (b_1^j - \beta_1^j)/s$, where $E(s^2) = \sigma^2$.

In terms of the original variates $\{X_i\}$

$$t = (b_1^j - \beta_1^j)/s \sqrt{\sum_{i=1}^{r} w_{i1}^2}$$

$i = 1, 2, \ldots, r$.

Next we need to determine $s^2$.

$$SSE = S(y - \sum b_1^iz_i)^2 = S\left[ (y - \sum \beta_1^iz_i) - (b_1^j - \beta_1^j)z_i \right]^2$$

$$= S(y - \sum \beta_1^iz_i)^2 - \sum (b_1^j - \beta_1^j)^2$$

$$= S(\bar{e} - \bar{e})^2 - \sum (b_1^j - \beta_1^j)^2$$

$$E(SSE) = (n - 1) \sigma^2 - r \sigma^2 = (n - r - 1) \sigma^2$$

Hence if we let $s^2 = SSE/(n-r-1)$, $E(s^2) = \sigma^2$.

If it is desired to make a test of the composite hypothesis $\{\beta_1^j = 0\}$ it would be useful to be able to use the $F$-ratio as a test criterion. In order to use $F$, it is necessary to find two quantities distributed as $F^{2} \sigma^2$, such that the ratio of the two will test the hypothesis $\{\beta_1^j = 0\}$ against the alternative $\{\beta_1^j \neq 0\}$. It would appear from the above that $SSE$ is distributed as $F^{2} \sigma^2$ with $(n-r-1)$ d.f. A more rigorous method of proof is the following:

Augment the existing set of $r$ completely orthogonal variates $\{z_i\}$ by $n-r-1$ others, which we shall designate as $\{p_j\}$ $(j = 1, 2, \ldots, n-r-1)$. The estimation equation will then be:

$$Y = \mu + \sum_{i=1}^{r} \beta_1^iz_i + e = \bar{Y} + \sum b_1^iz_i + \sum c_1^jp_j$$

where $c_1^j$ is the regression coefficient for $p_j(E \sum c_1^jp_j = 0)$. Note that with these $n$ orthogonal variates, there is no residual sum of squares. Hence

$$S\left[ (y - \sum b_1^iz_i - \sum c_1^jp_j)^2 = 0 \right]$$

$$0 = S\left[ (y - \sum \beta_1^iz_i - \sum (b_1^j - \beta_1^j)z_i - \sum c_1^jp_j)^2 \right]$$

$$= S(y - \sum \beta_1^iz_i)^2 - \sum (b_1^j - \beta_1^j)^2 - \sum (c_1^j)^2$$
\[ \sum_{j=1}^{n-r-1} (c_j^i)^2 = S(E - \hat{E})^2 - \sum_{i=1}^{r} (b_i^j - \beta_i^j)^2 = \text{SSE}. \]

Hence we have broken SSE into \((n - r - 1)\) orthogonal squares.

\[ c_j^i = S(y_{pj}) = S(E_{pj}) + \sum_i \beta_i^j S(z_{ij}) \]

\[ E\left[ c_j^i - \sum_i \beta_i^j S(z_{ij}) \right] = 0 \]

\[ E(c_j^i) = \sum_i \beta_i^j S(z_{ij}) = 0 \]

\[ \sigma^2(c_j^i) = E\left[ S(E_{pj}) \right]^2 = \sigma^2; \quad \sigma^2(c_j^i c_k^i) = 0. \quad (k \neq j) \]

Since the \{c_j^i\} are orthogonal linear forms in NID(0, \sigma^2) variates, they are NID(0, \sigma^2). Hence the \((c_j^i)^2\) are independently distributed as \(\chi^2 \sigma^2\) with one d.f. each or SSE is so distributed with \((n - r - 1)\) d.f. \(^1\)

We have already shown that the \((b_i^j - \beta_i^j)\) are NID(0, \sigma^2); hence, \((b_i^j - \beta_i^j)^2\) is distributed as \(\chi^2 \sigma^2\) with one d.f., and \(\sum (b_i^j)^2\) will be distributed as \(\chi^2 \sigma^2\) with \(r\) d.f. under the null hypothesis \(\{\beta_i^j = 0\}\). \(^1\)

Hence

\[ F = \frac{\sum (b_i^j)^2}{r} / \frac{\text{SSE}}{n - r - 1} \]

can be used to test the null hypothesis. Also we see that the single-tailed \(F\)-test should be used if the alternative hypothesis is \(\{\beta_i^j \neq 0\}\), because the expected value of the denominator is still \(\sigma^2\) while the expected value of the numerator is

\[ \frac{\sum E(b_i^j)^2}{r} = \sum \left[ E(b_i^j - \beta_i^j)^2 + 2E(\beta_i^j (b_i^j - \beta_i^j)) + E(\beta_i^j)^2 \right] / r \]

\[ = \sum \left[ \sigma^2 + (\beta_i^j)^2 \right] / r = \sigma^2 + \sum (\beta_i^j)^2 / r > \sigma^2. \]

\(^1\) Since the \{c_j^i\} and \{b_i^j\} are independent of one another, SSE and SSR are independent of one another. This is a necessary condition for the statements concerning \(F\).
There are many conditions under which it would be desirable to assume several of the $\beta_i \neq 0$. For a general test, we shall assume
\[
\left\{ \beta_1, \beta_2, \ldots, \beta_k \neq 0 \right\}
\]
and test the null hypothesis that
\[
\left\{ \beta_k + 1, \ldots, \beta_r = 0 \right\}
\]
We have shown that SSE, found by fitting $\left\{ \beta_1 \right\}$ for $i = 1, 2, \ldots, r$ is independent of the hypotheses for $\left\{ \beta_1 \right\}$ and that the $(\beta_1^2 - \beta_1^2)^2$ are independently distributed as $\chi^2 \sigma^2$ with one d.f. Hence under the null hypothesis that
\[
\sum_{i=k+1}^{r} (b_i^2)
\]
is distributed as $\chi^2 \sigma^2$ with $(r - k)$ d.f. We know that the reduction in the residual sum of squares due to fitting $\left\{ \beta_1 \right\}$ is given by
\[
\sum_{i=1}^{r} (b_i^2)
\]
and that due to fitting the first $k \left\{ b_i \right\}$ is
\[
\sum_{i=1}^{k} (b_i^2)
\]
Hence
\[
\sum_{i=k+1}^{r} (b_i^2)
\]
is the additional reduction in sum of squares gained by using
\[
\left\{ \beta_k + 1, \ldots, b_r \right\}
\]
in the regression equation after first using $\left\{ \beta_1, \ldots, b_k \right\}$. It should be emphasized that by use of the orthogonal transformation, every $b_i$ can be computed independently of all the others. Hence the reduction due to any one of the regression coefficients is independent of whether any others have been used and is independent of any assumptions made about the expected values of these others.
The major difficulty facing us at this stage is to transfer these results on the additional reduction due to the last \((r - k)\) variates back to the original set-up, where the value of any regression coefficient depends upon all the others included in the analysis. For example, the values of \(b_1, b_2, \ldots, b_k\) as estimated from all of the \(\{X_i\}\) will not be the same as the \(b'_1, b'_2, \ldots, b'_k\) as estimated from the first \(k\) \(\{X_i\}\). In order to use our orthogonal results, the following conditions must hold:

1. If we let \(R_r\) be the reduction in the residual sum of squares due to the first \(\{X_i\}\) and \(R_k\) that due to the first \(k\) \(\{X_i\}\) then \(\{R_r - R_k\}\) must be distributed as \(\chi^2\) with \((r - k)\) d.f., under the null hypothesis \[\beta_{k + 1} = \beta_{k + 2} = \cdots = \beta_r = 0\]

2. The residual sum of squares \(\Sigma y^2 - R_r\) is independently distributed as \(\chi^2 \sigma^2\) with \((n - r - 1)\) d.f.

We have shown that

\[\sum_{i = 1}^{k} (b'_i - \beta'_i)^2\]

is distributed as \(\chi^2\) with \(k\) d.f.,

\[\sum_{i = k + 1}^{r} (b'_i - \beta'_i)^2\]

with \((r - k)\) d.f., and

\[\text{SSE} = \sum_{i = 1}^{r} (b'_i - \beta'_i)^2 = \sum_{i = 1}^{r} (b'_i)^2 - \sum_{i = 1}^{r} (b'_i)^2\]

with \((n - r - 1)\) d.f. - all independent of one another.

But we have also shown that both methods of deriving the regression coefficients minimize \(S e^2\) and hence produce the same SSE. Hence

\[\Sigma y^2 - R_r = \text{SSE}\]

is distributed as \(\chi^2 \sigma^2\) with \((n - r - 1)\) d.f. Also
\[ R_r = \sum_{i=1}^{r} (b_i^2) \]

If we solve backwards for the \( \{ x_i \} \) in terms of the \( \{ z_i \} \) we find that
\[ x_i = w_{i1}^1 z_1 + w_{i2}^1 z_2 + \cdots + w_{i1}^1 z_1 \]
where \( w_{ij}^1 \) are functions of the \( w_{ij} \). For example
\[ w_{11}^1 = 1/w_{11} ; w_{21}^1 = -w_{21}/w_{11} w_{22} ; w_{22}^1 = 1/w_{22} \]

Since
\[ \sum \beta_j x_j = \sum \beta_j z_j \]
and by equating the coefficients of \( \{ z_i \} \), we find that
\[ \beta_j = \sum_{i=j}^{r} w_{ij} \beta_i \]

This shows that the null hypothesis \( \{ \beta_k + 1 \cdots , \beta_r = 0 \} \) means that
\[ \{ \beta_k + 1 \cdots , \beta_j = 0 \ \text{ because the } \beta_j \text{ are functions of only} \]
\[ \{ \beta_j , \beta_j + 1 \cdots , \beta_r \} \]. Hence
\[ \sum_{i=k+1}^{r} (b_i^1)^2 \]
is distributed as \( \chi^2 \sigma^2 \) with \( (r-k) \) d.f. under the null hypothesis for the \( \{ \beta_i \} \).

Now \( R_k \) is the reduction due to the first \( k \) \( \{ x_i \} \). But the first \( k \) \( \{ z_i \} \) are functions of only the first \( k \) \( \{ x_i \} \). For example
\[ z_k = \sum_{j=1}^{k} w_{kj} x_j \]

Again the reduction due to using the \( \{ x_i \} \) or \( \{ z_i \} \) must be the same, so that
\[ R_k = \sum_{i=1}^{k} (b_i^1)^2 \]
\[ R_r = \sum_{i=k+1}^{r} (b_i^1)^2 \]

the added reduction after the first \( k \) \( \{ x_i \} \) is given by
\[ R_r = R_k + \sum_{i=k+1}^{r} (b_i^1)^2 \]

It might be added that the \( \{ z_i \} \) behave just like orthogonal polynomials, which will be discussed in Section 8.6.
8.6. Curvilinear Regression: Orthogonal Polynomials.

8.6.1. Introduction. If it is desired to fit a regression equation using successive powers of one or more independent variates, the methods given in section 8.5 can be applied. For example, suppose we had a series of annual rainfall data \( Y \) for the years 1900-1949 and wished to determine a regression line such as the following polynomial:

\[
y = \alpha + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4 + \epsilon,
\]

where \( x = 1, 2, \cdots, 50 \). In order to use the methods of section 8.5, merely set \( x^1 = x_1 \). Snedecor (1) presents some examples using this technique of determining a polynomial regression line.

If the independent variable is equally spaced, such as in time or space, a convenient method of curve-fitting by orthogonal polynomials can be used.\(^1\) This method was first developed by Tchebycheff (22) and has since been extended by many writers. Some of the books and articles on this subject are (23), (24), (25), (26), (27), and (28). The advantage of orthogonal polynomials over the usual regression variates arises from the fact that the former are so constructed that any term of the polynomial is independent of any other term. This property of independence permits one to compute each regression coefficient independently of the others and also facilitates testing the significance of each coefficient. A summation method is illustrated by Fisher (3) and Snedecor (1), but the \( \hat{f} \) polynomial method presented by Fisher and Yates with tables for \( x \) through 52 (29) is generally more expeditious if tests of significance of the succeeding terms of the polynomial are desired. Another method described by Aitken (25) is also recommended if tables are available.

We shall consider the use of the \( \hat{f} \) polynomials, for which a computer's bulletin was prepared by Anderson and Housman, who also included an extension of

\(^1\)The method of orthogonal polynomials can be used for unequally spaced X's, but it is not so advantageous for such data.
the \( \sum \) tables through \( X = 104 \) (30). An example is given in this bulletin for 62 annual United States sugar prices, 1875 - 1936.

8.6.2. Determination of the polynomial. Suppose we wanted to fit a polynomial of high enough degree to \( n \) equally spaced points so that succeeding terms would not reduce the residual sum of squares by a significant amount, e.g.

\[
Y_j = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \beta_3 x_j^3 + \cdots + \beta r x_j^r + \varepsilon_j,
\]

where \( x_j = x_{\bar{j}} = j - \bar{j} \). Since \( \bar{j} = (n + 1)/2 \), we see that \( x_j \) takes on the values

\[-(n - 1)/2, -(n - 3)/2, \cdots, (n - 3)/2, (n - 1)/2.\]

Let us attempt to replace the above equation by one of the form

\[
Y_j = \alpha_0 + \alpha_1 P_{1j} + \alpha_2 P_{2j} + \cdots + \alpha_r P_{rj} + \varepsilon_j,
\]

where the \( \alpha \)'s are functions of the \( \beta \)'s and the \( P \)'s are functions of the powers of \( x_j \) chosen so that each \( P_{ij} \) is a function of all powers of \( x_j \) (\( = j - \bar{j} \)) equal to or less than \( i \):

\[
P_{ij} = C_{i0} + C_{i1} x_j + C_{i2} x_j^2 + \cdots + C_{i\ell} x_j^\ell.
\]

Let us construct a set of \( n \) polynomials, i.e., determine the coefficients, \( C \)'s, so that each polynomial is orthogonal to all others including \( P_0 = 1 \). That is,

\[
\sum P_{ij} P_{kj} = 0 \quad (i = 0, 1, \cdots, r; k \neq i).
\]

In the future, \( S \) will refer to summation over \( j \). For a given \( P_{ij} \), we want the \( i \) sums for \( k = 0, 1, \cdots, (i - 1) \) to be zero. Since each polynomial is a function of the powers of \( x_j \) equal to or less than \( k \), we can replace the above summations by

\[
\sum P_{ij} x_j^k = 0 \quad (k = 0, 1, 2, \cdots, i - 1).
\]

That is, we have \( i \) equations to determine the values of the \((i + 1)\) coefficients:

\[
C_{i0}, C_{i1}, \cdots, C_{i\ell}.
\]

Hence it is necessary to fix the value of one constant; we shall set \( C_{i\ell} = 1 \).

---

2/ We are using the notation \( P \) instead of \( \sum \) for ease of typing.
Using the definition of $P_{ij}$ in terms of the $x$'s, we have for a given $k$

$$S(C_{i0} + C_{i1}x_j + \cdots + C_{i,i-1}x_j^{i-1} + x_j^i) x_j^k = \sum_{i=0}^{i-1} C_{i,i-1}x_j^{i-1}$$

We know that $Sx^m = 0$ when $m$ is odd. Hence we have the following equations for $i$ odd:

<table>
<thead>
<tr>
<th>$k$</th>
<th>equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$C_{i,0} + C_{i,2}x_2 + \cdots + C_{i,i-1}x_2^{i-1} = 0$</td>
</tr>
<tr>
<td>2</td>
<td>$C_{i,0}x_2^2 + C_{i,2}x_2^4 + \cdots + C_{i,i-1}x_2^{i+1} = 0$</td>
</tr>
<tr>
<td>i-1</td>
<td>$C_{i,0}x_2^{i-1} + C_{i,2}x_2^{i+1} + \cdots + C_{i,i-1}x_2^{2(i-1)} = 0$</td>
</tr>
<tr>
<td>1</td>
<td>$C_{i,1}x_2 + C_{i,3}x_2^3 + \cdots + C_{i,i-2}x_2^{i-1} + x_2^{i+1} = 0$</td>
</tr>
<tr>
<td>3</td>
<td>$C_{i,1}x_2^3 + C_{i,3}x_2^5 + \cdots + C_{i,i-2}x_2^{i+1} + x_2^{i+3} = 0$</td>
</tr>
<tr>
<td>i-2</td>
<td>$C_{i,1}x_2^{i-1} + C_{i,3}x_2^{i+1} + \cdots + C_{i,i-2}x_2^{2i-4} + x_2^{2i-2} = 0$</td>
</tr>
</tbody>
</table>

The $\frac{1}{2}(i+1)$ equations for $k$ even involve only the coefficients $\{C_{i,0}, C_{i,2}, \cdots, C_{i,i-1}\}$, and since these equations have no constant term, all the coefficients are zero. But the $\frac{1}{2}(i-1)$ equations for $k$ odd do have non-zero constant terms, and we can solve for the coefficients

$$\left\{C_{i,1}, C_{i,3}, \cdots, C_{i,i-2}\right\}$$

Hence we conclude that when $i$ is odd ($i = 1, 3, \cdots$),

$$P_1 = x$$

$$P_3 = x^3 + C_{3,1}x$$

$\cdots$

$$P_i = x^i + C_{i,1}x^{i-2} + \cdots + C_{i,3}x^3 + C_{i,1}x$$

$\cdots$
Similarly it can be shown that when \( i \) is even (\( i = 0, 2, 4 \))

\[
\begin{align*}
P_0 &= 1 \\
P_2 &= x^2 + C_2,0 \\
P_4 &= x^4 + C_4,2x^2 + C_4,0 \\
&
\ddots
\end{align*}
\]

\[
P_i = x^i + C_{i,i-2}x^{i-2} + \cdots + C_2x^2 + C_2,0
\]

Note that we have omitted the \( j \) subscript for convenience.

From these results we see that

\[
x^\ell P_i = P_i+1 + D_{1,1}P_{i-1} + D_{1,2}P_{i-3} + \cdots,
\]

where the \( D \)'s are functions of the \( C \)'s. For example

\[
x^\ell P_1 = x^2 = P_2 - C_2,0
\]

\[
x^\ell P_2 = x^3 + C_2,0x = P_3 + (C_2,0 - C_3,1)x
\]

\[
D_{11} = -C_2,0; \quad D_{21} = C_2,0 - C_3,1; \quad D_{1,\ell} = 0 \quad \text{for} \quad \ell > 1.
\]

Since \( (x^\ell P_1) \) is a function of powers of \( x \) equal to or less than \( (i + 1) \), the orthogonality requirement, \( \mathbb{S}_\ell^\ell x^k = 0 \) for \( \ell > k \), guarantees that

\[
\mathbb{S}_\ell x^\ell P_1 = 0
\]

for \( \ell > (i + 1) \) and because of the symmetry of \( \ell \) and \( i \) for \( i > (\ell + 1) \). Also since \( P_i^2 \) is a function of even powers of \( x \), \( \mathbb{S}_\ell P_i^2 = 0 \). Hence the only cases for which \( \mathbb{S}_\ell x^\ell P_1 \neq 0 \) are \( \ell = i \pm 1 \). And when \( \ell = i \pm 1 \), all of the terms except those for \( P_{i+1} \) and \( P_{i-1} \) vanish. That is

\[
\mathbb{S}_\ell x^\ell P_{i+1} = \mathbb{S}_\ell^2 P_{i+1} \quad \text{or} \quad \mathbb{S}_\ell x^\ell P_{i-1} = \mathbb{S}_\ell^2 P_i
\]

\[
\mathbb{S}_\ell x^\ell P_{i-1} = D_{1,1} x^\ell P_{i-1}
\]

Hence

\[
D_{1,1} = \frac{\mathbb{S}_\ell^2 P_i}{\mathbb{S}_\ell^2 P_{i-1}}
\]

and all other \( D \)'s = 0.

Therefore we now have a recursion formula to determine higher-order polynomials in terms of lower-order ones:
\[ P_1 + 1 = xP_1 - \frac{SP_1^2}{SP_1 - 1} P_1 - 1 \]

Since \( P_1 = x \), \( P_0 = 1 \), \( SP_1^2 = Sx^2 = n(n^2 - 1)/12 \) and \( SP_0^2 = n \),

\[ P_2 = x^2 - \frac{n^2 - 1}{12} \]

Similarly

\[ SP_2^2 = Sx^4 - n \left( \frac{n^2 - 1}{12} \right) = \frac{n(n^2 - 1)(n^2 - 4)}{180} \]

and

\[ P_3 = x^3 - \left( \frac{n^2 - 1}{12} + \frac{n^2 - 4}{15} \right) x = x^3 - \frac{3n^2 - 7}{20} x \]

A general formula for \( SP_1^2 \) is

\[ SP_1^2 = \frac{n(n^2 - 1)(n^2 - 4) \cdots (n^2 - i^2)(1!)^2 \left( \frac{i - 1}{i} \right)^2}{4(2i + 1) \left( (2i - 1)! \right)^2} \]

Hence

\[ D_{1,1} = \frac{SP_1^2}{SP_1^2 - 1} = \frac{(n^2 - i^2)i^2}{4(4i^2 - 1)} \]

A proof of these results is beyond the scope of this book but can be found in most of the sources quoted in section 8.6.1. Hence

\[ P_4 = x^4 - \left[ \frac{3n^2 - 7}{20} + \frac{(n^2 - 9)9}{140} \right] x^2 + \frac{9(n^2 - 1)(n^2 - 9)}{(12)(140)} \]

\[ = x^4 - \frac{3n^2 - 13}{14} x^2 + \frac{3(n^2 - 1)(n^2 - 9)}{560} \]

The reader will note that these polynomial values, \( P_{ij} \), will be fractions. For convenience of computing and of presentation of tables of the polynomial values, it is desirable to use only integral values for the \( P_{ij} \). The regular polynomial values have been multiplied by a constant, \( \lambda_1 \), in (29) and (30), so chosen that the values of \( P_{ij}^1 = \lambda_1 P_{ij} \) are integers reduced to lowest terms.
Hence

\[ Y_j = \alpha_0 + \alpha_1 p_{1j} + \alpha_2 p_{2j} + \cdots + \alpha_r p_{rj} + e_j, \]

where

\[ \alpha_i = \lambda_i, \quad \text{and} \quad \alpha_0 = \lambda_0. \]

Example 8.6.1. As an example, consider the construction of the \( P_{ij} \) values for \( n = 5 \) (refer to example 6.4 on page 76). Obviously, for 5 points, a fourth degree polynomial will fit the data exactly; hence, only \( P_0 \) (i.e., \( P_1, P_2, P_3, \) and \( P_4 \) exist. Using the formulas already indicated, we have \( x_j = j - \frac{n+1}{2} = j - 3, \)
and

\[ P_{1j} = x_j; \quad P_{2j} = x_j^2 - 2; \quad P_{3j} = x_j^3 - \frac{17}{5} x_j; \quad P_{4j} = x_j^4 - \frac{31}{7} x_j^2 + \frac{72}{35}. \]

The polynomial values are

\[
\begin{array}{cccccccc}
\text{\( j \)} & \text{\( P_{1j} \)} & \text{\( P_{2j} \)} & \text{\( P_{3j} \)} & \text{\( P_{4j} \)} & \text{\( P_{3j} \)} & \text{\( P_{4j} \)} \\
1 & -2 & 2 & 6/5 & 12/35 & 1 & 1 \\
2 & -1 & -1 & 12/5 & -4/35 & 2 & -4 \\
3 & 0 & -2 & 0 & 72/35 & 0 & 6 \\
4 & 1/2 & -1 & -12/5 & -4/35 & -2 & 4 \\
5 & 2 & -2 & -6/5 & 12/35 & -1 & 1 \\
\end{array}
\]

\[
S(P_1')^2 = \frac{10}{2} = 10 \quad \text{and} \quad 70
\]

We note that \( P_{1j} = P_1' \) and \( P_{2j} = P_2' \); \( \lambda_1 = \lambda_2 = 1, \lambda_3 = 5/6 \) and \( \lambda_4 = 35/12, \) so that \( P_{3j} = 5P_{3j}/6 \) and \( P_{4j} = 35P_{4j}/12. \)

8.6.3. Estimating the parameters, \( \alpha_1 \) and \( \sigma^2. \) We are given the regression equation to fit to \( n \) equally spaced points:

\[ Y_j = \sum_{i=0}^{r} \alpha_i p_{ij} + e_j = \hat{Y}_j + e_j, \quad (r + 1 \leq n), \]

where the \( \{e_j\} \) are assumed NID \((0, \sigma^2)\) and

\[ \hat{Y}_j = \alpha_0 + \alpha_1 p_{1j} + \cdots + \alpha_r p_{rj}, \]

with \( \alpha_1 = \lambda_1. \) The least squares equations to determine the \( \lambda_i \) are quite simple,
because of the orthogonal property of the $P_i$'s. In fact $A_0 = \bar{Y}$ and

$$A_1^i = \frac{S(YP_i^1)}{S(P_i^1)^2}.$$

The polynomial tables in (30) present values of the $P_i$'s, $\lambda_i$, and $S(P_i^1)^2$ for values of $\lambda$ through 104. In these tables, $P_i^1 = \int_1^i$.

The sum of squares of deviations from the regression is given by

$$SSR = n\bar{Y}^2 + A_1^1 S(YP_1^1) + A_2^1 S(YP_2^1) + \cdots + A_r^1 S(YP_r^1),$$

where each term is the independent reduction due to the successive polynomials; means, $P_1^1, P_2^1, \cdots, P_r^1$. Also

$$SSE = SY^2 - SSR = SY^2 - \sum_{i=1}^{r} SSR_i,$$

where

$$SSR_i = A_i^1 S(YP_i^1) = (SYP_i^1)^2 / S(P_i^1)^2.$$

$s^2 = SSE/(n - r - 1)$ is an unbiased estimate of $\sigma^2$. Hence we can set up the following Analysis of Variance

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear regression ($P_1^1$)</td>
<td>1</td>
<td>$SSR_1 = (SYP_1^1)^2 / S(P_1^1)^2$</td>
</tr>
<tr>
<td>Quadratic regression ($P_2^1$)</td>
<td>1</td>
<td>$SSR_2 = (SYP_2^1)^2 / S(P_2^1)^2$</td>
</tr>
<tr>
<td>n$^{th}$ degree regression ($P_r^1$)</td>
<td>1</td>
<td>$SSR_r = (SYP_r^1)^2 / S(P_r^1)^2$</td>
</tr>
<tr>
<td>Error</td>
<td>$n - r - 1$</td>
<td>$s^2 = SSE/(n - r - 1)$</td>
</tr>
</tbody>
</table>

In general we do not know what degree ($r$) should be used; so tests of successive terms are made until there is no material reduction in $s^2$ by the use of additional terms.

It should be pointed out that the equation

$$\hat{Y} = A_0^1 + A_1^1 P_1^1 + \cdots + A_r^1 P_r^1$$

can be used for prediction purposes. However if the original polynomial form of the equation is needed, the $P_i^1$ will have to be replaced by their equivalent
polynomial function,

\[ P_i^j = \sum_{k=0}^{n} \lambda_k (c_{10} + c_{11}x + c_{12}x^2 + \cdots + x^k) \]

This is adequately explained in (30).

Example 8.6.2. In order to compare the number of spears per asparagus plant for male and female plants, the California Agricultural Experiment Station at Davis, obtained the following data on the differences (Y) between the mean number of spears per plant for each of the years 1925-1936, here indicated as years 1-12.

<table>
<thead>
<tr>
<th>Year (j)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>Sum of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>1.1</td>
<td>7.1</td>
<td>11.0</td>
<td>12.6</td>
<td>14.7</td>
<td>19.9</td>
<td>25.1</td>
<td>23.9</td>
<td>23.1</td>
<td>23.6</td>
<td>26.0</td>
<td>24.6</td>
<td>4516.43</td>
</tr>
<tr>
<td>P_1</td>
<td>-11</td>
<td>-9</td>
<td>-7</td>
<td>-5</td>
<td>-3</td>
<td>-1</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>572</td>
</tr>
<tr>
<td>P_2</td>
<td>55</td>
<td>25</td>
<td>1</td>
<td>-17</td>
<td>-29</td>
<td>-35</td>
<td>-29</td>
<td>-17</td>
<td>1</td>
<td>25</td>
<td>55</td>
<td></td>
<td>12012</td>
</tr>
<tr>
<td>P_3</td>
<td>-33</td>
<td>3</td>
<td>21</td>
<td>25</td>
<td>19</td>
<td>7</td>
<td>-7</td>
<td>-19</td>
<td>-25</td>
<td>-21</td>
<td>-3</td>
<td>33</td>
<td>5148</td>
</tr>
<tr>
<td>P_4</td>
<td>33</td>
<td>-27</td>
<td>-33</td>
<td>-13</td>
<td>12</td>
<td>28</td>
<td>28</td>
<td>12</td>
<td>-13</td>
<td>-33</td>
<td>-27</td>
<td>33</td>
<td>8008</td>
</tr>
</tbody>
</table>

The first four sets of polynomial values are given below the yields. The \( \lambda \)'s are: \( \lambda_1 = 2, \ \lambda_2 = 3, \ \lambda_3 = 2/3, \ \lambda_4 = 7/24. \)

\[
\begin{align*}
SY &= 212.7 \\
SP_{1Y} &= 602.4 \\
SP_{2Y} &= -1025.7 \\
SP_{3Y} &= -19.5 \\
SP_{4Y} &= 71.7 \\
A_0' &= 17.725 \\
A_1' &= 1.05262 \\
A_2' &= -0.058296 \\
A_3' &= -0.00378788 \\
A_4' &= 0.00395355 \\
nS &= 3770.11 \\
SSR_1 &= 633.78 \\
SSR_2 &= 87.58 \\
SSR_3 &= 0.07 \\
SSR_4 &= 0.64
\end{align*}
\]

The analysis of variance is as follows:
<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>d.f.</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon^2$</td>
<td>11</td>
<td>746.32</td>
<td></td>
</tr>
<tr>
<td>Linear regression ($P_1^1$)</td>
<td>1</td>
<td>633.78</td>
<td>633.78**</td>
</tr>
<tr>
<td>Deviation from linear</td>
<td>10</td>
<td>112.54</td>
<td>11.25</td>
</tr>
<tr>
<td>Quadratic regression ($P_2^1$)</td>
<td>1</td>
<td>87.58</td>
<td>87.58**</td>
</tr>
<tr>
<td>Deviation from quadratic</td>
<td>9</td>
<td>24.96</td>
<td>2.77</td>
</tr>
<tr>
<td>Cubic regression ($P_3^1$)</td>
<td>1</td>
<td>.07</td>
<td>0.07</td>
</tr>
<tr>
<td>Deviation from cubic</td>
<td>8</td>
<td>24.89</td>
<td>3.11</td>
</tr>
<tr>
<td>Cubic and quartic regression</td>
<td>2</td>
<td>.71</td>
<td>0.36</td>
</tr>
<tr>
<td>($P_3^1, P_4^1$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deviation from quartic</td>
<td>7</td>
<td>24.25</td>
<td>3.46</td>
</tr>
</tbody>
</table>

On the basis of these results, we conclude that the regression is quadratic and that the best estimate of $\sigma^2$ is 2.77 with 9 degrees of freedom. It should be indicated that successive terms of the polynomial are tested by use of the deviation from regression. For example, we test for the existence of linear regression by use of:

$$F' = \frac{(633.78)/(11.25)}{1} = 56.34$$

with (1, 10) degrees of freedom. This is not an exact test, because we find out later that the denominator is an over-estimate of $\sigma^2$. However, the test is certainly on the "safe side"; that is, if $F'$ is significant, the $F$ by use of the best estimate of $\sigma^2$ will certainly be significant. Hence we can use $F'$ as a guide to the advisability of considering any form of regression. It is always advisable to compute at least one more polynomial value after the first non-significant one. That is we might have stopped with $P_3^1$, but it is advisable to try out $P_4^1$ to make sure it is not significant.

Using the quadratic regression equation,

$$\hat{Y} = 17.725 + 1.053P_1^1 - 0.0854P_2^1$$

$$= 17.725 + 2.106(j - 6.5) - 0.0854 \left[3(j - 6.5)^2 - 35.75\right]$$

$$= 17.725 + 2.106(j - 6.5) - 0.0854(3j^2 - 39j + 91)$$

$$= -3.735 + 5.437j - 0.2562j^2$$

$$= (1.45, 6.11, 10.27, 13.91, 17.04, 19.66, 21.77, 23.36, 24.45, 25.02, 25.07, 24.61)$$
Exercise 8.6.1. (a) Prove that \( A_i^1 \) is an unbiased estimate of \( \chi_i^1 \)
\[
\sigma^2(A_i^1) = \sigma^2 / S(P_i^1)^2, \quad \text{and} \quad \sigma^2(A_i^1 A_i^1) = 0 \quad \text{for } i \neq k.
\]
(b) Prove that \( s^2 \) is an unbiased estimate of \( \sigma^2 \).
(c) Prove that SSR is independent of SSE when \( \chi_i^1 = 0 \).

Exercise 8.6.2. Derive the formula for \( P_5 \).

Exercise 8.6.3. Compute the polynomial values, \( P_i^1 \) (\( i = 0, 1, \ldots, 5 \)), for \( n = 6 \).

Exercise 8.6.4. (a) Determine the standard errors of \( A_i^1 \) and \( A_i^2 \) in Example 8.6.2 and then set up 95% confidence limits for the true regression coefficients \( \chi_i^1 \) and \( \chi_i^2 \).
(b) Do the same for \( A_1 \) and \( A_2 \) as estimates of \( \chi_1 \) and \( \chi_2 \).
(c) What is the standard error of \( \hat{Y}_i \)?

Exercise 8.6.5. Given the following data on the output (\( Y_1 \)) in lbs./manhr. for General Motors employees and the aggregate weight of all General Motors auto products in millions of lbs. (\( Y_2 \)) for the years 1929-1941:

<table>
<thead>
<tr>
<th>Year</th>
<th>1929</th>
<th>1930</th>
<th>1931</th>
<th>1932</th>
<th>1933</th>
<th>1934</th>
<th>1935</th>
<th>1936</th>
<th>1937</th>
<th>1938</th>
<th>1939</th>
<th>1940</th>
<th>1941</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y_1 )</td>
<td>12.55</td>
<td>13.41</td>
<td>14.79</td>
<td>14.56</td>
<td>15.05</td>
<td>15.67</td>
<td>17.60</td>
<td>18.30</td>
<td>18.61</td>
<td>19.69</td>
<td>19.87</td>
<td>20.81</td>
<td>21.83</td>
</tr>
<tr>
<td>( Y_2 )</td>
<td>5632</td>
<td>3537</td>
<td>3175</td>
<td>1692</td>
<td>2477</td>
<td>3812</td>
<td>5338</td>
<td>6617</td>
<td>6868</td>
<td>4020</td>
<td>5618</td>
<td>7492</td>
<td>8631</td>
</tr>
</tbody>
</table>

(a) Use a set of orthogonal polynomial tables (29) or (30) or construct such a set for \( n = 13 \) to determine the best fitting polynomial to fit to each set of data.
(b) If the same degree equation is to be used for both sets of data, what degree would you advise?
(c) How would you determine the regression of \( Y_2 \) on \( Y_1 \), both adjusted for time trends? What theoretical difficulties do you see in the determination of the regression of \( Y_2 \) on \( Y_1 \)?
Exercise 8.6.6. Snedecor (1) presents the following problem on the average weights of sunflowers:

Week 1 2 3 4 5 6 7 8 9 10 11 12
Height 18 36 68 98 131 170 206 228 247 250 254 254

(a) What degree of polynomial should be used to estimate the height after \( j \) weeks?
(b) Plot the actual and estimated heights.
(c) Determine the standard errors of the regression coefficients.

Exercise 8.6.7. We know that the graph of the equation

\[ Y = 10 - 2X + 3X^2 - X^3 \]

passes through the points

\( (1, 10), (2, 10), (3, 4), (4, -14), (5, -50), (6, -110) \).

Use these points to derive the prediction equation

\[ \hat{Y} = A_0 + A_1P_1 + A_2P_2 + A_3P_3 \]

and show that \( \hat{Y} = Y \) in this case.

Exercise 8.6.8. The student is advised to select some data from his own field of research for which a polynomial prediction equation is useful and to carry out the necessary computations to determine \( \hat{Y} \).
8.7. The Regression Problem when Certain Assumptions are Relaxed.

Suppose we relax the assumptions of normality and homoscedasticity in the previous sections of this chapter. For this case, David and Neyman (31) give a proof of the following theorem on Least Squares. Given

(i) \( Y_1, Y_2, \ldots, Y_n \) are independent.
(ii) \( E(Y_j) = \sum_{i=1}^{r} \beta_i X_{ij} \) and the \( X \)'s are fixed variates.\(^1\)
(iii) Out of the \( n \) equations in (ii), it is possible to select at least one system of \( r \) equations soluble with respect to the \( \beta \)'s.
(iv) The variances of the \( Y_j \) satisfy the relationship

\[
\sigma_j^2 = \sigma^2 w_j \quad (j = 1, 2, \ldots, n),
\]

where \( \sigma^2 \) may be unknown but the \( w_j \) are known positive constants \( > 0 \).

Then

(a) The best linear estimate of \( Y \) is

\[
\hat{Y} = \sum b_i X_i
\]

where the \( b \)'s are obtained by minimizing the weighted sum of squares

\[
SSE_w = \sum_{j=1}^{r} w_j (Y_j - \sum_{i=1}^{r} b_i X_{ij})^2
\]

with respect to the \( b_i \). The typical least squares equation is

\[
\frac{\partial SSE_w}{\partial b_k} = 0; \quad b_1 SwX_1X_k + \cdots + b_k SwX_k^2 + \cdots + b_r SwX_rX_k = SwX_kY.
\]

(b) The estimate of \( \sigma^2(\hat{Y}) \) is given by

\[
s^2(\hat{Y}) = \frac{-\Delta_0 \Delta_1}{(n - r) \Delta^2},
\]

\(^1\) If it is desired to use the intercept \( \alpha \) in the equation, set \( X_{1j} = 1 \).
where

\[
\Delta_0 = \begin{bmatrix}
H_0 & H_1 & \cdots & H_r \\
H_1 & G_{11} & \cdots & G_{1r} \\
\vdots & \vdots & & \vdots \\
H_r & G_{r1} & \cdots & G_{rr}
\end{bmatrix}, \quad \Delta_1 = \begin{bmatrix}
0 & X_1 & \cdots & X_r \\
X_1 & G_{11} & \cdots & G_{1r} \\
\vdots & \vdots & & \vdots \\
X_r & G_{r1} & \cdots & G_{rr}
\end{bmatrix}
\]

\[
\Delta = \begin{bmatrix}
G_{11} & \cdots & G_{1r} \\
\vdots & & \vdots \\
G_{r1} & \cdots & G_{rr}
\end{bmatrix}
\]

\[
H_0 = SwY^2 \\
H_i = SwX_iY \\
G_{ii} = SwX_iX_i
\]

If \( r = 2 \), with \( \beta_1 = \alpha \) and \( \beta_2 = \beta \), we have

\[
\hat{Y} = a + bX = \bar{Y}_w + b(X - \bar{X}_w),
\]

where

\[
X_1 \equiv 1, \ X_2 \equiv X, \ \bar{Y}_w = SwY/Sw \text{ and } \bar{X}_w = SwX/Sw.
\]

Also \( b = Swxy/Swx^2 \), where \( x = X - \bar{X}_w \) and \( y = Y - \bar{Y}_w \). \( \bar{X}_w \) and \( \bar{Y}_w \) are called weighted means and \( b \) a weighted regression coefficient. Often the weights are adjusted so that \( Sw = 1 \).

**Example 8.7.1.** The following data are taken from an experiment on soybeans with 3 nitrogen treatments plus a check treatment (no nitrogen) to find out the relation between the mean yield of soybeans per acre (\( Y \)) and amount of nitrogen in lbs. per acre (\( X \)).

<table>
<thead>
<tr>
<th>lbs. N/acre (X)</th>
<th>0</th>
<th>47.8</th>
<th>94</th>
<th>157</th>
</tr>
</thead>
<tbody>
<tr>
<td>bu. beans/acre (Y)</td>
<td>14.7</td>
<td>14.6</td>
<td>17.8</td>
<td>22.1</td>
</tr>
<tr>
<td>no. of plots</td>
<td>14</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>( w )</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Since 14 plots were used for the 0 level of nitrogen and 7 plots for the 3 treatments, so weights of (2, 1, 1, 1) were used in estimating the regression of mean yield on amount of nitrogen per acre. For the three treatments, the variance of the mean yield for the 0 nitrogen plots was one-half that of the mean yields for the 3 treatments.
The calculations needed were:

\[ \bar{Y}_W = 59.60 \quad \bar{Y}_W = 16.78 \]
\[ S_{Wx}^2 = 17,933.20 \quad S_{Wxy} = 828.66 \quad S_{Wy}^2 = 42.75 \]
\[ b = 0.0462 \]
\[ \hat{Y} = 16.78 + 0.0462 (X - 59.60) . \]

In this case the estimate of variance was taken from the \( 13 + 3(6) = 31 \) degrees of freedom within treatments: \( s^2 = 7.56 \). Hence the estimated variance of each of the \( Y \)'s, which were means of \( 7w \) plots, was \( 7.56/7w = 1.08/w \).

**Exercise 8.7.1.** (a) Show that \( s^2(\hat{Y}) \) for \( r = 2 \), using the David-Neyman result, is

\[ s^2(\hat{Y}) = \frac{S_{Wy}^2 - (S_{Wxy})^2}{n - 2} \left( \frac{1}{S_Y^2} + \frac{x^2}{S_{Wx}^2} \right) . \]

(b) Show that for Example 8.7.1, \( s^2(\hat{Y}) = 2.22(\frac{1}{S_Y} + \frac{x^2}{S_{Wx}^2}) \).

**Exercise 8.7.2.** (a) In Example 8.7.1, show that the estimated variance of \( b \) is

\[ s^2(b) = \frac{1.08}{S_{Wx}^2} = \frac{1.08}{17,933} \]

by use of the formula for the variance of a linear form.

(b) Hence show that the variance of \( \hat{Y} \) is

\[ s^2(\hat{Y}) = 1.08 \left( \frac{1}{S_Y} + \frac{x^2}{S_{Wx}^2} \right) = 0.216 + x^2 \]

(c) Can you explain why this estimate of \( \sigma^2(\hat{Y}) \) is smaller than that given using the David-Neyman formula (Exercise 8.7.1)?

**Exercise 8.7.3.** G. A. Baker (32) furnishes an example on ovarian weights (in milligrams) of rats receiving five dosages of Serum VII, 16 rats for each dosage.

<table>
<thead>
<tr>
<th>Dosage in rat units (X)</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>12</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean wt. (Y)</td>
<td>27.00</td>
<td>38.31</td>
<td>43.44</td>
<td>65.25</td>
<td>72.88</td>
</tr>
<tr>
<td>( s^2(Y) )</td>
<td>1.59</td>
<td>3.96</td>
<td>3.64</td>
<td>18.96</td>
<td>45.95</td>
</tr>
<tr>
<td>Smoothed weights (w)</td>
<td>7.72</td>
<td>2.16</td>
<td>1.00</td>
<td>0.37</td>
<td>0.19</td>
</tr>
</tbody>
</table>

The "smoothed weights" were found by smoothing the values of \( s(Y) \).
(a) Show that if we use the "smoothed weights"

\[ \hat{Y} = -10.15 + 62.37 \log_{10} X, \]

where \( \log_{10} X \) is used as the independent variable.

(b) Use as weights reciprocals of \( s^2(Y) \) to derive another estimate of \( \hat{Y} \).

(c) Would you adjudge the differences in the regression lines of (a) and (b) to be significant?

REFERENCES CITED


3. Bartlett, M. S. "Fitting a Straight Line when Both Variables are Subject to Error." *Biometrics* 5:207-212 (1949).


OTHER READING


CHAPTER 9
LEAST SQUARES FOR EXPERIMENTAL DESIGN MODELS

2.1. Introduction.

The least squares methods outlined before can be readily applied to the
problem of testing the significance of the differences among a set of means. We
have shown that the t-test is a valid test for testing if two samples could have
come from the same \( N(\mu, \sigma^2) \). A computational device which furnishes the
requisite statistics for making a test for many samples (two being a special case)
is the analysis of variance. The analysis of variance is a simple arithmetic device
for dividing the total sum of squares into separate independent parts. The analysis
of variance was actually introduced in the previous sections, when we set up the
following general analysis for the regression equation:

\[ \bar{Y} = \mu + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=k+1}^{r} \beta_i x_i + \epsilon \]

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>First ( k ) variates</td>
<td>( k )</td>
<td>( R_k )</td>
<td></td>
</tr>
<tr>
<td>Added reduction by last ( (r-k) ) variates</td>
<td>( r-k )</td>
<td>( R_T - R_k )</td>
<td>((R_T - R_k)/(r-k))</td>
</tr>
<tr>
<td>Error</td>
<td>( n-r-1 )</td>
<td>( SSE )</td>
<td>( s^2 = SSE/(n-r-1) )</td>
</tr>
</tbody>
</table>

We can test the null hypothesis that the last \( (r-k) \) \( \beta \)'s are 0, without
assuming anything about the first \( k \) \( \beta \)'s, by use of \( F = (R_T - R_k)/(r-k)s^2 \),
with \( (r-k) \) and \( (n-r-1) \) degrees of freedom. We generally wish to eliminate
the effect of the mean from the regression first and do not want to test if the
mean is 0. Hence we have indicated everything as deviations from the mean with
\( \mu \) estimated by \( \bar{Y} \). When it is desired to test the null hypothesis that any
sub-set of \( (r-k) \) of the \( \beta \)'s are all zero, we would determine the reduction,
\( R_T \), due to all \( r \) variates and then the reduction \( R_k \) due to the other \( k \) variates
alone. The difference between the two reductions furnishes the necessary data to
test the desired hypothesis by use of $F$, as indicated above. If the two sets of variates happen to be independent, then we could also test the first $k$ variates using $F = R_k / k s^2$. This merely means that the added reduction for these $k$ variates after fitting the last $(r - k)$ variates alone is the same as $R_k$, or conversely that the reduction due to fitting the last $(r - k)$ variates alone is the same as $R_r - R_k$.

In many experiments it is possible to divide the $X$ variates into independent sets of variates, so that the sums of the individual reductions in the total variation add to the total reduction, such as with orthogonal polynomials. We shall consider a number of different experimental models, many of which have this desirable characteristic. One of the reasons for designing an experiment with independent sets of $X$ variates is to cut down on the computations needed to determine the $b$'s and the reductions in the total variation attributable to these $b$'s. For example, if we have 10 regression coefficients to estimate, we must invert a $10 \times 10$ matrix, which is a rather lengthy task. However, if we have 2 sets of 5 variates, we have to invert two $5 \times 5$ matrices, a much shorter task.

In the analysis of experimental data, we shall generally consider the following steps:

1. Construct a regression model which expresses symbolically the nature of the variables used in the experiment and the method of conducting the experiment.

2. Determine estimates of the parameters in the regression model by least squares.

3. Indicate the reduction in the total variation due to the regression in an analysis of variance table.

4. Compute $s^2$ and the necessary $F$-values.

5. Compute variances of the estimates in (2) and confidence limits.

6. If possible, compare the error variance using this design with the expected variance if any other design had been used – this is called the efficiency of the design.

7. Examples.
We shall consider several types of experimental models in the succeeding sections. Many of the techniques of analysis were originally presented to the authors by W. G. Cochran. The reader is encouraged to read Chapter 3 of (1) for a more complete account of these techniques. Some of the other books and articles on this subject are also cited at the end of this chapter.

9.2. Completely Randomized Design.

9.2.1. Suppose we have \( p \) different treatments which are to be tested for yield producing ability. This yield can be achievement scores in education, length of life of a commercial product, bushels per acre of a crop, profit from a given economic plan, etc. Let us assume that each treatment is planted on \( r \) different experimental units which we shall call plots. Then we might estimate the yield for the \( i^{th} \) treatment on the \( j^{th} \) plot of its \( r \) plots as:

\[
Y_{ij} = \mu + \sum_{k=1}^{p} \tau_k x_{ki} + \epsilon_{ij},
\]

where

\[
x_{ki} = \begin{cases} 
0 & \text{for } k \neq i \\
1 & \text{for } k = i 
\end{cases}; \quad i = 1, 2, \ldots, p; \quad j = 1, 2, \ldots, r,
\]

\( \tau_i \) is the differential effect of the \( i^{th} \) treatment (over the mean, \( \mu \)) and

\[
\sum_{i} \sum_{k} \tau_k x_{ki} = 0. \quad \text{Some restriction such as the last one is necessary, since there will be only} \ p \ \text{treatment totals to estimate} \ (p + 1) \ \text{constants} - \mu \ \text{and the} \ p \ \tau \ \text{values}; \ \text{and if the particular one mentioned above (} \sum \sum \tau_k x_{ki} = 0 \) \ \text{is used,} \ \mu \ \text{will be the over-all mean effect.}
\]

Since the \( x_i \)'s are 0 or 1, the regression model can be simplified to

\[
Y_{ij} = \mu + \tau_i + \epsilon_{ij},
\]

where \( \sum \tau_i = 0. \) As usual it is assumed that the \( \epsilon_{ij} \) are NID(0, \( \sigma^2 \)) and independent of the \( \tau_i \). Hence the expected mean yield for the \( i^{th} \) treatment is \( \mu + \tau_i \).
We wish to estimate the values of $\mu$ and $\overline{\epsilon}_1$ and $\sigma^2$ from an experiment which has produced $r$ yields for each of the $p$ treatments. The method of experimentation is to select $rp$ plots in the field and then to allocate $p$ treatments at random on these plots, with the restriction that each treatment should appear on $r$ plots. Let us designate the estimates of $\mu$ and $\overline{\epsilon}_1$ as $m$ and $t_1$. Hence the regression equation can be set up in the form

$$Y_{ij} = m + t_i + e_{ij},$$

where $e_{ij}$ is the estimate of $e_{ij}$. $m$ and the $\{t_i\}$ are to be chosen so as to minimize $S_e^2$.

The least squares equations are:

$$m = \frac{rpm + \sum t_i}{r} = \frac{SY}{G},$$
$$t_i = \frac{rm + r^2t_i}{\sum Y_{ij}} = T_i$$

where $T_i$ is the total yield for the $i^{th}$ treatment and $G$ the grand total for the experiment. If the $t_i$ equations are added together, this sum is exactly $m$ equation. This shows that the $p$ treatment equations are not independent of the equation for $m$. In order to solve these equations, an auxiliary relationship must be used. Since $\sum \overline{\epsilon}_1 = 0$, a reasonable relationship is

$$\sum t_i = 0,$$

so that each $t_i$ measures a differential effect from the mean. Also this makes $m = G/rp = \overline{Y}$, the general mean. And $t_i = \frac{T_i}{r} - m = (T_i - \overline{T})/r$, where $\overline{T} = \frac{G}{p}$.

If the average yield for the $i^{th}$ treatment is desired, we calculate

$$t_i = t_i + m = \frac{T_i}{r}.$$

Using the regression model and the least squares equation for $G$, it is seen that

$$G = rpm + r \sum t_i + r \sum e_{ij} = rpm + r \sum t_i.$$

Hence $\sum \sum e_{ij} = 0$. That is, the sum of the estimated residuals equals zero.

Or we may consider an experiment with $p$ treatments and $r$ random samples for each treatment.
2.2.3. The total reduction in sum of squares due to regression is $\sum b_1 x_i y = \sum b_1 x i y$. In the case of treatments, this reduction is

$$\sum_{i=1}^{p} t_i (T_i - \bar{T}) = \sum_{i=1}^{p} (T_i - \bar{T})^2 / r = SST.$$

We compute $SST = \frac{\sum T_i^2}{r} - C$, where $C = g^2 / rp$. Expressed in terms of the original model,

$$T_i - \bar{T} = (r / \alpha + r \bar{x}_i / j + \sum j \epsilon_{ij} - (rp / \alpha + r \sum i \bar{l}_i + \sum i \sum j \epsilon_{ij}) / p$$

$$= r(\bar{T}_i - \bar{T}) + \frac{(p - 1)}{p} \sum j \epsilon_{ij} - \frac{1}{p} \sum_\ell \neq i \sum j \epsilon_{ij}$$

$$E(T_i - \bar{T}) = r(\bar{T}_i - \bar{T}).$$

But it has been shown that $\sum \bar{T}_i = p \bar{T} = 0$ is a necessary part of the original model. Therefore

$$E(T_i - \bar{T})^2 = r^2 \bar{T}_i^2 + \left[ \frac{(p - 1)}{p} + \frac{1}{p} \right] r(p - 1)$$

$$= r^2 \bar{T}_i^2 + \frac{r(p - 1)}{p} \sigma^2,$$

and

$$E(SST) = r \sum_{i=1}^{p} T_i^2 = (p - 1) \sigma^2.$$

The analysis of variance table is

<table>
<thead>
<tr>
<th></th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>Mean Square</th>
<th>E(MS)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>$p - 1$</td>
<td>SST</td>
<td>MST = SST / $(p - 1)$</td>
<td>$r \sigma^2_t + \sigma^2$</td>
</tr>
<tr>
<td>Error</td>
<td>$p(r - 1)$</td>
<td>SSE</td>
<td>$s^2 = SSE / p(r - 1)$</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>

*E(MS) = Expected value of the mean square; $\sigma^2_t = \sum \bar{T}_i^2 / (p - 1)$. 
From Chapter 8, we know that SSE = \( Sy^2 - SST \), where SST is the reduction due to regression (in this case the treatments), and SSE has \( n - p = rp - p = p(r - 1) \) degrees of freedom. If we set \( s^2 = SSE/p(r - 1) \), \( Es^2 = \sigma^2 \). Hence we can use \( F = MST/s^2 \) to test \( H_0: \{ \tau_i = 0 \} \) against the alternative that some \( \tau_i \neq 0 \). The one-tailed F-test is used since the numerator of \( F \) is expected to be greater than the denominator when any \( \tau_i \neq 0 \).

Since \( t_i^1 = T_i/r = (r \mu + r \tau_i + \sum_j \varepsilon_{i,j})/r \) and \( m = G/rp = \mu + \bar{\varepsilon} \), it is seen that

\[
E(t_i^1) = \mu + \bar{\tau}_i; \quad En = \mu \\
\sigma^2(t_i^1) = \sigma^2/r; \quad \sigma^2(m) = \sigma^2/rp
\]

Also for any other treatment \( t_j^1 = T_j/r \),

\[
\sigma^2(t_i^1t_j^1) = 0
\]

The estimated difference between the mean effects of these two treatments is

\[
d = t_i^1 - t_j^1 = t_i^1 - t_j^1
\]

\[
Ed = \bar{\tau}_i - \bar{\tau}_j
\]

Since \( t_i^1 \) and \( t_j^1 \) are independent, \( \sigma^2(d) = 2 \sigma^2/r \) and the confidence limits for \( \sigma \) are

\[
d - t_\alpha s \sqrt{2/r} < d < d + t_\alpha s \sqrt{2/r}
\]

Example 9.2.1. As a first example, we might consider an experiment which was run to determine the number of warp skin breaks on tent twill in 5 consecutive days of testing with 4 breaks per test (2). The results are

<table>
<thead>
<tr>
<th>Day</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30</td>
<td>30</td>
<td>40</td>
<td>45</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>40</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>45</td>
<td>55</td>
<td>35</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>45</td>
<td>35</td>
<td>40</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>130</td>
<td>160</td>
<td>175</td>
<td>160</td>
<td>210</td>
<td>835</td>
</tr>
<tr>
<td>Mean</td>
<td>32.5</td>
<td>40.0</td>
<td>43.75</td>
<td>40.0</td>
<td>52.5</td>
<td>41.75</td>
</tr>
</tbody>
</table>
Analysis of Variance

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days</td>
<td>4</td>
<td>845.00</td>
<td>211.25*</td>
</tr>
<tr>
<td>Error</td>
<td>15</td>
<td>668.75</td>
<td>44.58</td>
</tr>
</tbody>
</table>

The error variance is estimated to be 44.58. There are significant differences among the day means, since \( F = 211.25/44.58 = 4.74 \) with 4 and 15 degrees of freedom.

Also the standard error of the difference between any two day means is given by

\[
s(d) = \sqrt{2(44.58)/4} = 4.72.
\]

Hence the 95% confidence limits for the difference between the mean number of breaks for the first two days are

\[
7.5 - 10.1 < \bar{d} < 7.5 + 10.1 \quad \text{or} \quad -2.6 < \bar{d} < 17.6,
\]

where \( 10.1 = t_{0.025} = (2.13)(4.72), \) with 15 degrees of freedom.

Example 9.2.2. A second example presents an analysis of the gains in weight (grams/100 days) of rats on a stock ration with various amounts of Gossypol added (Halvorson and Sherwood, N. C. State College, 1932).

<table>
<thead>
<tr>
<th>Rats</th>
<th>None</th>
<th>.04%</th>
<th>.07%</th>
<th>.10%</th>
<th>.13%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>228</td>
<td>186</td>
<td>179</td>
<td>130</td>
<td>154</td>
</tr>
<tr>
<td>2</td>
<td>229</td>
<td>228</td>
<td>193</td>
<td>87</td>
<td>130</td>
</tr>
<tr>
<td>3</td>
<td>218</td>
<td>220</td>
<td>183</td>
<td>135</td>
<td>130</td>
</tr>
<tr>
<td>4</td>
<td>216</td>
<td>208</td>
<td>180</td>
<td>116</td>
<td>118</td>
</tr>
<tr>
<td>5</td>
<td>222</td>
<td>228</td>
<td>143</td>
<td>118</td>
<td>118</td>
</tr>
<tr>
<td>6</td>
<td>208</td>
<td>198</td>
<td>204</td>
<td>165</td>
<td>104</td>
</tr>
<tr>
<td>7</td>
<td>235</td>
<td>222</td>
<td>114</td>
<td>151</td>
<td>112</td>
</tr>
<tr>
<td>8</td>
<td>229</td>
<td>273</td>
<td>188</td>
<td>59</td>
<td>134</td>
</tr>
<tr>
<td>9</td>
<td>233</td>
<td>216</td>
<td>178</td>
<td>126</td>
<td>98</td>
</tr>
<tr>
<td>10</td>
<td>219</td>
<td>198</td>
<td>134</td>
<td>64</td>
<td>100</td>
</tr>
<tr>
<td>11</td>
<td>224</td>
<td>213</td>
<td>208</td>
<td>78</td>
<td>104</td>
</tr>
<tr>
<td>12</td>
<td>220</td>
<td>196</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>232</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>200</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>208</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>232</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Sum  | 3555 | 2391 | 2100 | 2043 | 1302 |
| Mean | 222.2| 217.4| 175.0| 120.2| 118.4|

[1]
\[ C = \frac{(11391)^2}{67} = 1,936,640. \]

**Sum of Squares**

Total: \((228)^2 + (229)^2 + \ldots + (178)^2 - C = 178,985.\]

Rations: \(\frac{(3555)^2}{16} + \frac{(2391)^2}{11} + \frac{(2100)^2}{12} + \frac{(2043)^2}{17} + \frac{(1302)^2}{11} - C = 140,083.\]

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rations</td>
<td>4</td>
<td>140,083</td>
<td>35,020.75**</td>
</tr>
<tr>
<td>Error</td>
<td>62</td>
<td>38,902</td>
<td>627.45</td>
</tr>
</tbody>
</table>

The only complication is that the number of samples per ration is not the same for all rations. In this experiment, \(s^2 = 627.45\); also \(F = 55.8\), a highly significant value (\(F < .01\)).

**Exercise 9.2.1.** (a) In Example 9.2.1, use the method of orthogonal polynomials, outlined in section 8.6 to divide the Sum of Squares for Days (SSD) into four orthogonal parts, each with one degree of freedom (linear, quadratic, cubic, and quartic). For example the polynomial values \(P^1_i\) for the linear effect are \((-2, -1, 0, 1, 2)\). Consider the totals as the \(Y\) values and show that \(SP^1_i = 160\). It has been proven that \(SSR_1 = (SP^1_i)^2 / S(P^1_i)^2 = 25,600 / 10 = 2,560\), but this is larger than SSD. Under the null hypothesis of no day differences, show that \(E(\text{SSR}_1) = 4 \cdot 2\). Hence to put SSR1 on the same basis as \(s^2\) (whose expected value is \(2\)), we must divide SSR1 by \(r = 4\), so that the linear effect is \(2560 / 4 = 640\).

(b) Complete the analysis of the quadratic, cubic, and quartic effects. Is there a significant departure from linearity?

(c) Suppose we had used the means as the \(Y\)-values. What changes must be made in the above analyses?
(d) Can you make any general statement of how to analyze by orthogonal polynomials equally spaced sets of observations with observations at each point?

Exercise 9.2.2. (a) Set up the mathematical model for Example 9.2.2. Solve the least squares equations and show that the analysis presented there is correct.

(b) Prove that the standard error of the difference between the effects of any two treatments is \( s \sqrt{\frac{1}{r_1} + \frac{1}{r_2}} \), where \( r_1 \) and \( r_2 \) are the numbers of rats on the two rations. Hence show that the standard error of the difference between the gains for rations No Gossypol and .04% is 9.81.

(c) What is the standard error of the average of the first two against the third ration?

Exercise 9.2.3. An investigation of North Carolina farmers' retail produce markets was made by John Curtis of the Department of Agricultural Economics, N. C. State College (1948). Data were collected on the dollar value of livestock owned by a sample of sellers on these markets, as follows:

<table>
<thead>
<tr>
<th>Seller</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>721</td>
<td>750</td>
<td>3480</td>
<td>627</td>
<td>469</td>
<td>812</td>
<td>393</td>
<td>369</td>
<td>332</td>
<td>249</td>
<td>1106</td>
<td>1703</td>
<td>1462</td>
</tr>
<tr>
<td>2</td>
<td>64</td>
<td>756</td>
<td>293</td>
<td>169</td>
<td>604</td>
<td>271</td>
<td>841</td>
<td>785</td>
<td>842</td>
<td>371</td>
<td>1702</td>
<td>563</td>
<td>1088</td>
</tr>
<tr>
<td>3</td>
<td>664</td>
<td>1315</td>
<td>370</td>
<td>976</td>
<td>165</td>
<td>1100</td>
<td>426</td>
<td>655</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>134</td>
<td>293</td>
<td>284</td>
<td>109</td>
<td></td>
<td>305</td>
<td>1947</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>610</td>
<td>865</td>
<td>119</td>
<td>704</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>546</td>
<td>3274</td>
<td>1980</td>
<td>2275</td>
<td>704</td>
<td>359</td>
<td>725</td>
<td>2195</td>
<td>97</td>
<td>428</td>
<td>422</td>
<td>295</td>
<td>73</td>
</tr>
<tr>
<td>7</td>
<td>278</td>
<td></td>
<td>697</td>
<td></td>
<td></td>
<td></td>
<td>477</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3046</td>
<td>7253</td>
<td>6526</td>
<td>4839</td>
<td>1238</td>
<td>2488</td>
<td>1660</td>
<td>3756</td>
<td>1174</td>
<td>3904</td>
<td>4924</td>
<td>4427</td>
<td>7309</td>
</tr>
<tr>
<td>No. Sellers</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Mean</td>
<td>381</td>
<td>1209</td>
<td>1088</td>
<td>691</td>
<td>413</td>
<td>622</td>
<td>553</td>
<td>939</td>
<td>587</td>
<td>558</td>
<td>1231</td>
<td>738</td>
<td>312</td>
</tr>
</tbody>
</table>

Grand totals

<table>
<thead>
<tr>
<th></th>
<th>52544</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>761.5</td>
</tr>
</tbody>
</table>
(a) Analyze these data to see if there are any real differences among markets.

(b) Compare the mean value per seller for the first nine markets with the mean value for markets 10 and 11.

Exercise 2.2.4. Snedecor and Cox (3) analyzed some data on the gain in weight (grams) of 149 Wistar rats over a six weeks' period for four successive generations. The gains for the males and females in each of these generations were:

<table>
<thead>
<tr>
<th>Generation</th>
<th>Malo</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>No. of rats</td>
<td>21</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>Total gain</td>
<td>3716</td>
<td>2422</td>
<td>1868</td>
</tr>
<tr>
<td>Mean gain</td>
<td>176.95</td>
<td>161.47</td>
<td>155.67</td>
</tr>
</tbody>
</table>

The total sum of squares adjusted for the mean was $SS^2 = 176.836$.

(a) Set up the analysis of variance among and within the 8 classes and determine if there are significant differences among the 8 mean gains.

(b) What is the standard error to test for the difference between the mean gain for the males of a given generation and the mean gain for the females of the same generation? What would you say regarding the differences between the population means for each of the 4 generations?

(c) Can you make a statement regarding the difference between males and females over all 4 generations?

2.3 Randomized Complete Blocks.

2.3.1 Suppose that $rp$ plots are divided into $r$ blocks of $p$ plots each and that each treatment is assigned at random to one of these plots in each block. Then the estimation equation is

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} = m + t_i + b_j + e_{ij}$$

where $\beta_j$ is the differential effect of the $j$th block ($j = 1, 2, \cdots, r$) being estimated by $b_j$, and $\sum_j \beta_j = 0$. 
2.3.2. The least squares equations are
\[
    m : \text{rpm} + r \sum t_i + p \sum b_j = G
\]
\[
    t_i : r(m + t_i) + \sum b_j = T_i
\]
\[
    b_j : p(m + b_j) + \sum t_i = B_j,
\]

where \( B_j \) is the total yield for the \( j \)th block. In this case, two auxiliary equations are required since \( \sum T_i = \sum \beta_j = 0 \) and in order to make \( m = \bar{T} = G/rp \), we set \( \sum t_i = \sum b_j = 0 \). Hence \( t_i = (T_i - \bar{T})/r \) and \( b_j = (B_j - \bar{B})/p \), where \( \bar{T} = G/p \) and \( \bar{B} = G/r \). We will let
\[
    t'_i = t_i + m = T_i/r; \quad b'_j = b_j + m = B_j/p.
\]

2.3.3. The total reduction due to blocks and treatments is
\[
    \frac{\sum (T_i - \bar{T})^2}{r} + \frac{\sum (B_j - \bar{B})^2}{p} = \text{SST} + \text{SSB}.
\]

We note that
\[
    rt_i = T_i - \bar{T} = r T_i - \sum \epsilon_{ij} - \sum \sum \epsilon_{ij}/p
\]
\[
    pb_j = B_j - \bar{B} = p \beta_j + \sum \epsilon_{ij} - \sum \sum \epsilon_{ij}/p.
\]

Hence \( Et_i = T_i \) and \( Eb_j = \beta_j \). Also \( \left( \sum (T_i - \bar{T})(B_j - \bar{B}) \right) = (1 - 1 - 1 + 1) \sigma^2 = 0 \) and SST and SSB are independent parts of the total reduction. The analysis of variance is

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degrees of freedom</th>
<th>Sum of squares</th>
<th>Mean square</th>
<th>E(MS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>( r - 1 )</td>
<td>SSB</td>
<td>MSB</td>
<td>( \sigma^2 + p \sigma^2_b )</td>
</tr>
<tr>
<td>Treatments</td>
<td>( p - 1 )</td>
<td>SST</td>
<td>MST</td>
<td>( \sigma^2 + r \sigma^2_t )</td>
</tr>
<tr>
<td>Error</td>
<td>((r - 1)(p - 1))</td>
<td>SSE</td>
<td>( s^2 )</td>
<td>( \sigma^2 )</td>
</tr>
</tbody>
</table>

where \( \sigma^2_b = \sum \beta_j^2/(r - 1) \).
9.3.6. $s^2 = \text{SSE}/(r - 1)(p - 1)$ and $F = \text{MST}/s^2$ to test for over-all treatment differences. Note that $\text{SSE} = S_y^2 - \text{SST} - \text{SSB}$. There is usually no desire to test for block differences, since the blocks are generally chosen to be different.

9.3.5. As in 9.2.5, $\sigma^2 (t') = \sigma^2 /r$ and $\sigma^2 (d) = 2 \sigma^2 /r$.

9.3.6. We might investigate the efficiency of this experimental design in which the treatments have been planted so that every treatment appears in every block as compared to a completely randomized design (9.2). Let us designate the error variance for the completely randomized design as $\sigma^2_w$ and for the randomized complete blocks design as $\sigma^2_{rb}$. We note that the expected total sum of squares, $\text{ESy}^2$, for the completely randomized design is

$$(p - 1)(r \sigma^2_t + \sigma^2_w) + p(r - 1) \sigma^2_w$$

and for the randomized complete blocks

$$(r - 1)(p \sigma^2_b + \sigma^2_{rb}) + (p - 1)(r \sigma^2_t + \sigma^2_{rb}) + (r - 1)(p - 1) \sigma^2_{rb}.$$ 

Assuming that $\sigma^2_t$ and $\sigma^2_b$ would not have been changed if we had used a completely randomized design instead of a randomized complete blocks design, we conclude that the expected total sum of squares would be the same in the two cases. Hence

$$(p - 1) \sigma^2_w + p(r - 1) \sigma^2_w = (r - 1)(p \sigma^2_b + \sigma^2_{rb}) + (p - 1) \sigma^2_{rb} + (r - 1)(p - 1) \sigma^2_{rb},$$

$$\sigma^2_w = \frac{r(p - 1) \sigma^2_{rb} + (r - 1)(p \sigma^2_b + \sigma^2_{rb})}{rp - 1}$$

$$s^2_w = \frac{r(p - 1) s^2_{rb} + \text{SSB}}{rp - 1} = s^2_{w2}$$

The ratio of $s^2_w$ to $s^2_{rb}$ is the estimated efficiency (I) of the randomized blocks design as compared to the completely randomized design. 2/

$$I = s^2_w/s^2_{rb}.$$ 

2/ We shall use the symbol I for efficiency, since we have already used E for expectation.
Example 9.3.1. This example presents the analysis of the differences between fat absorption by doughnuts in 8 different fats, all being tested on each of 6 days.\(^3\)

<table>
<thead>
<tr>
<th>Grams of Fat Absorbed by Mixes of 24 Doughnuts During Cooking Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fat Number</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>01</td>
</tr>
<tr>
<td>02</td>
</tr>
<tr>
<td>03</td>
</tr>
<tr>
<td>04</td>
</tr>
<tr>
<td>05</td>
</tr>
<tr>
<td>06</td>
</tr>
<tr>
<td>Totals</td>
</tr>
<tr>
<td>Means</td>
</tr>
</tbody>
</table>

\[ G = \frac{(8292)^2}{48} = 1,432,443 \]

\[ Sy^2 = (164)^2 + (172)^2 + \cdots + (167)^2 - 1,432,443 = 9,143.00 \]

\[ SSF = \frac{(1032)^2 + (1067)^2 + \cdots + (974)^2}{6} - 1,432,443 = 3,344.33 \]

\[ SSD = \frac{(1331)^2 + (1472)^2 + \cdots + (1479)^2}{8} - 1,432,443 = 3,986.75 \]

\[ SSE = Sy^2 - SSF - SSD = 1,811.92 \]

### Analysis of Variance

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between day means</td>
<td>5</td>
<td>3,986.75</td>
<td>797.35</td>
</tr>
<tr>
<td>Between fat means</td>
<td>7</td>
<td>3,344.33</td>
<td>477.76**</td>
</tr>
<tr>
<td>Error</td>
<td>35</td>
<td>1,811.92</td>
<td>51.77 = s^2</td>
</tr>
</tbody>
</table>

**Significant at 1 percent probability level.**

We note that the efficiency of this design as compared to the completely randomized design is

\[ I = \frac{42(51.77) + 3,986.75}{47(51.77)} = \frac{6161.09}{2433.19} = 2.53 \]

This indicates that if the fats had been distributed randomly over the six days, the error variance would have been expected to be about 2.53 as large as that with the randomized blocks design, where each fat was used each day. In other words,

In this example F stands for treatments and D(f) for blocks.
we have made our estimates of fat differences about \(2^{1/2}\) times as precise by planning the experiment so that every fat was used on each day of the experiment. The 95% confidence limits for the difference between any two fat means are 
\[ d - 8.43 < \bar{d} < d + 8.43 \]. A brief word of caution about the use of these confidence limits should be presented. Those are average confidence limits, assuming that you select the treatments to be compared in advance of the experiment.

If you wait until the experiment is over and then select, for example, the highest and lowest treatment means to compare, the confidence limits are much wider than 
\[ \bar{d} \pm t \frac{s(d)}{\sqrt{n}} \]. One excellent article on this point is presented by J. W. Tukey (4). The reader might also refer to reference (5) on "order statistics".

**Exercise 9.3.1.** Read reference (4).

**Exercise 9.3.2.** In the investigation of North Carolina farmers' markets by Curtis (Exercise 9.2.3), the incomes of a sample of regular, former, and potential patrons were studied at 10 markets with these results:

<table>
<thead>
<tr>
<th>City</th>
<th>Regular</th>
<th>Former</th>
<th>Potential</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asheville*</td>
<td>30</td>
<td>30</td>
<td>28</td>
<td>88</td>
</tr>
<tr>
<td>Asheboro</td>
<td>39</td>
<td>36</td>
<td>27</td>
<td>102</td>
</tr>
<tr>
<td>Charlotte*</td>
<td>41</td>
<td>37</td>
<td>28</td>
<td>106</td>
</tr>
<tr>
<td>Durham*</td>
<td>29</td>
<td>29</td>
<td>24</td>
<td>82</td>
</tr>
<tr>
<td>Greensboro*</td>
<td>32</td>
<td>27</td>
<td>22</td>
<td>81</td>
</tr>
<tr>
<td>Raleigh*</td>
<td>29</td>
<td>27</td>
<td>29</td>
<td>85</td>
</tr>
<tr>
<td>Goldsboro</td>
<td>25</td>
<td>24</td>
<td>20</td>
<td>69</td>
</tr>
<tr>
<td>Rocky Mount</td>
<td>27</td>
<td>28</td>
<td>27</td>
<td>82</td>
</tr>
<tr>
<td>Franklin</td>
<td>30</td>
<td>29</td>
<td>24</td>
<td>83</td>
</tr>
<tr>
<td>Jacksonville</td>
<td>29</td>
<td>28</td>
<td>27</td>
<td>84</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>311</td>
<td>295</td>
<td>256</td>
<td>862</td>
</tr>
</tbody>
</table>

*Cities of over 50,000 population
(a) Would you conclude that there are any real income differences among the three groups of people?

(b) What are the confidence limits for the difference between the average income of the regular and potential patrons?

(c) Is there any feature of this analysis which might be open to question?

Exercise 9.3.3. An analysis was made of the fiber diameters in microns at 6 different regions on the seed coat of the Mexican 128 variety of cotton (6). The analysis was made on a sample of 10 seeds as follows:

<table>
<thead>
<tr>
<th>Plant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>16.49</td>
<td>17.80</td>
<td>17.54</td>
<td>16.75</td>
<td>17.54</td>
<td>17.54</td>
<td>103.66</td>
</tr>
<tr>
<td>B</td>
<td>15.45</td>
<td>15.96</td>
<td>15.71</td>
<td>14.13</td>
<td>14.40</td>
<td>14.40</td>
<td>90.05</td>
</tr>
<tr>
<td>C</td>
<td>16.23</td>
<td>15.96</td>
<td>16.49</td>
<td>14.92</td>
<td>14.66</td>
<td>14.92</td>
<td>93.18</td>
</tr>
<tr>
<td>D</td>
<td>18.33</td>
<td>17.28</td>
<td>16.49</td>
<td>16.49</td>
<td>17.28</td>
<td>17.80</td>
<td>103.67</td>
</tr>
<tr>
<td>E</td>
<td>16.49</td>
<td>18.33</td>
<td>17.54</td>
<td>17.02</td>
<td>17.28</td>
<td>18.06</td>
<td>104.72</td>
</tr>
<tr>
<td>F</td>
<td>16.49</td>
<td>17.54</td>
<td>17.05</td>
<td>15.71</td>
<td>15.45</td>
<td>14.66</td>
<td>96.90</td>
</tr>
<tr>
<td>G</td>
<td>15.96</td>
<td>15.71</td>
<td>16.23</td>
<td>16.49</td>
<td>15.18</td>
<td>16.49</td>
<td>96.06</td>
</tr>
<tr>
<td>H</td>
<td>16.75</td>
<td>16.23</td>
<td>14.66</td>
<td>15.96</td>
<td>13.35</td>
<td>16.75</td>
<td>93.70</td>
</tr>
<tr>
<td>I</td>
<td>14.40</td>
<td>18.33</td>
<td>17.02</td>
<td>14.66</td>
<td>15.71</td>
<td>17.02</td>
<td>97.14</td>
</tr>
<tr>
<td>J</td>
<td>16.49</td>
<td>17.02</td>
<td>16.75</td>
<td>17.54</td>
<td>15.71</td>
<td>16.49</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Total 163.08 170.16 165.48 159.67 156.56 164.13 979.08

(a) Show that the analysis of variance is correct and fill in the degrees of freedom.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plants</td>
<td></td>
<td>4.14</td>
</tr>
<tr>
<td>Regions</td>
<td></td>
<td>2.22</td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td>.582</td>
</tr>
</tbody>
</table>

(b) What statement would you make regarding region differences?

(c) What is the standard error of a region mean?

(d) What are the confidence limits for the difference between regions 1 and 6?

Exercise 9.3.4. Middleton and Chapman conducted an experiment to determine the best variety out of eight varieties of oats at Laurinburg, North Carolina, in 1940. The yields of grain in grams for a 16' row were:

...
Blocks

<table>
<thead>
<tr>
<th>Variety</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>Sum</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>296</td>
<td>357</td>
<td>340</td>
<td>331</td>
<td>348</td>
<td>1672</td>
<td>334.4</td>
</tr>
<tr>
<td>2</td>
<td>402</td>
<td>390</td>
<td>431</td>
<td>340</td>
<td>320</td>
<td>1883</td>
<td>376.6</td>
</tr>
<tr>
<td>3</td>
<td>437</td>
<td>334</td>
<td>426</td>
<td>320</td>
<td>296</td>
<td>1813</td>
<td>362.6</td>
</tr>
<tr>
<td>4</td>
<td>303</td>
<td>319</td>
<td>310</td>
<td>260</td>
<td>242</td>
<td>1434</td>
<td>286.8</td>
</tr>
<tr>
<td>5</td>
<td>469</td>
<td>405</td>
<td>442</td>
<td>487</td>
<td>394</td>
<td>2197</td>
<td>439.4</td>
</tr>
<tr>
<td>6</td>
<td>345</td>
<td>342</td>
<td>358</td>
<td>300</td>
<td>308</td>
<td>1653</td>
<td>330.6</td>
</tr>
<tr>
<td>7</td>
<td>324</td>
<td>339</td>
<td>357</td>
<td>352</td>
<td>220</td>
<td>1592</td>
<td>318.4</td>
</tr>
<tr>
<td>8</td>
<td>488</td>
<td>374</td>
<td>401</td>
<td>338</td>
<td>320</td>
<td>1921</td>
<td>384.2</td>
</tr>
<tr>
<td>Sum</td>
<td>3064</td>
<td>2860</td>
<td>3065</td>
<td>2728</td>
<td>2448</td>
<td>14165</td>
<td>354.1</td>
</tr>
</tbody>
</table>

(a) Determine if there are significant differences among the varieties. Which variety would you recommend?

(b) What is the efficiency of this design compared to a completely randomized design?

(c) What is the standard error for the difference between two varietal means?

**Exercise 9.3.5.** Frequently one or more plots in a randomized blocks experiment will be missing because of some adverse condition such as a wash-out or insect infestation. As a result the orthogonal properties of the analysis are upset, and the statistician must either use an approximate analysis or run a complete least squares equation. Let us assume in the oats experiment of Exercise 9.3.4 that the plot for variety 1 in block I had been washed out. An approximate method of analysis is to set this yield equal to $\bar{y}$, run up the analysis of variance in terms of $\bar{y}$ and the other yields and estimate $\bar{y}$ by minimizing SSE. Then substitute $\bar{y}$ for the missing plot yield and complete the analysis, decreasing the error degrees of freedom by one.

(a) Estimate $\bar{y}$ by the method outlined above and compute the analysis of variance.

(b) Make a complete least squares analysis of the data, regarding the missing plot as non-existent; in other words, we have only 4 plots for variety 1 and 7 plots in block I. Show that you obtain the same value of SSE as in (a) but that SST in (a) is too large by $(2768 - 7\bar{y})^2/56$. 
9.4. Latin Square Designs.

A design slightly more complicated than the randomized blocks design is the Latin square design. With this design each treatment is assigned at random within a row and a column so that all treatments appear in each row and column. Hence it is possible to adjust the error for plot heterogeneity in two directions. Of course the rows may be fields and the columns similar locations in these fields. The basic design looks as follows for 3 treatments (there then must be 3 rows and 3 columns).

<table>
<thead>
<tr>
<th></th>
<th>C₁</th>
<th>C₂</th>
<th>C₃</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>R₁</td>
<td>T₁(23)</td>
<td>T₂(17)</td>
<td>T₃(29)</td>
<td>69</td>
</tr>
<tr>
<td>R₂</td>
<td>T₂(16)</td>
<td>T₃(25)</td>
<td>T₁(16)</td>
<td>57</td>
</tr>
<tr>
<td>R₃</td>
<td>T₃(24)</td>
<td>T₁(18)</td>
<td>T₂(12)</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td></td>
<td>180</td>
</tr>
</tbody>
</table>

The figures represent yields in an experiment. In the field the rows and columns of the basic design would have been randomized. One such field arrangement would be

<table>
<thead>
<tr>
<th></th>
<th>C₁</th>
<th>C₂</th>
<th>C₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>R₁</td>
<td>T₃</td>
<td>T₁</td>
<td>T₂</td>
</tr>
<tr>
<td>R₂</td>
<td>T₂</td>
<td>T₃</td>
<td>T₁</td>
</tr>
<tr>
<td>R₃</td>
<td>T₁</td>
<td>T₂</td>
<td>T₃</td>
</tr>
</tbody>
</table>

The regression model for this design is:

\[ Y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \varepsilon_{ijk} \]

where \( \alpha_i \) and \( \beta_j \) are row and column effects and \( \tau_k \) the treatment effect.

We note that once \( i \) and \( j \) are specified in a particular field arrangement, we know \( k \). Hence \( k \) is a function of \( i \) and \( j \). For example in the field arrangement mentioned above if \( i = 2 \) and \( j = 3 \), \( k = 1 \).
The analysis of variance table is similar to that of the randomized complete blocks design except that there are two sets of blocks — rows and columns. The analysis of the 3 x 3 example is as follows:

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows</td>
<td>2</td>
<td>42</td>
<td>21</td>
</tr>
<tr>
<td>Columns</td>
<td>2</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Treatments</td>
<td>2</td>
<td>186</td>
<td>93</td>
</tr>
<tr>
<td>Error</td>
<td>2</td>
<td>6</td>
<td>3</td>
</tr>
</tbody>
</table>

Example 9.5.1. In the mixing of an explosive powder, four mixes were used. (2). The powder was charged into cartridge primers by a number of operators, four of whom were selected for the test, which was carried on for four days. The mixer-charger combination was changed each day, each combination being used once. Test results on the product gave the following (chargers are indicated by numbers 1 to 4, mixers by letters W to Z):

<table>
<thead>
<tr>
<th>Day</th>
<th>Charger</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>W</td>
<td>71</td>
<td>75</td>
<td>54</td>
<td>44</td>
<td>244</td>
</tr>
<tr>
<td></td>
<td>Z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuesday</td>
<td>X</td>
<td>62</td>
<td>64</td>
<td>64</td>
<td>54</td>
<td>244</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wednesday</td>
<td>Y</td>
<td>60</td>
<td>65</td>
<td>57</td>
<td>45</td>
<td>227</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Z</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>W</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thursday</td>
<td>Z</td>
<td>65</td>
<td>47</td>
<td>44</td>
<td>69</td>
<td>225</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>W</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>258</td>
<td>251</td>
<td>219</td>
<td>212</td>
<td>940</td>
</tr>
</tbody>
</table>

The analysis of variance is:

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days</td>
<td>3</td>
<td>81.50</td>
<td>27.17</td>
</tr>
<tr>
<td>Chargers</td>
<td>3</td>
<td>196.25</td>
<td>65.42</td>
</tr>
<tr>
<td>Mixers</td>
<td>3</td>
<td>125.25</td>
<td>41.75</td>
</tr>
<tr>
<td>Error</td>
<td>6</td>
<td>1072</td>
<td>178.67</td>
</tr>
</tbody>
</table>

All F values are less than 1.
Exercise 9.4.1. (a) The student should set up the least squares equations and the analysis of variance table for an \( r \times r \) Latin square design and indicate the expected values of the mean squares in the analysis of variance.

(b) What is the standard error of the difference between any two treatment means?

(c) Show that the efficiency of the rows in reducing the error variance in an \( r \times r \) Latin square design as compared to a randomized complete blocks design with the \( r \) columns as blocks is given by

\[
I = \frac{SSR + (r - 1)^2 s^2_L}{r(r - 1)s^2_L},
\]

where SSR is the sum of squares for rows and \( s^2_L \) is the error mean square for the Latin square design. Similarly the efficiency of the columns is found by replacing SSR by SSC (sum of squares for columns). Hint: Use the same method as given on page 186 for the efficiency of the randomized complete blocks design as compared to a completely randomized design.

Exercise 9.4.2. Given a Latin square design with \( r \) rows, columns, and treatments. Assume that the data for one plot were lost. Show that the estimate of the missing value found by minimizing SSE is

\[
y = \frac{r(R + C + T) - 2G}{(r - 1)(r - 2)},
\]

where \( R, C, \) and \( T \) are respectively the totals of the row, column, and treatment which contain the missing value.

Exercise 9.4.3. A Latin square design was used at the University of Hawaii to compare 6 different legume intercycle crops for pineapples. The yields in 10 gram units are given below.
Table:

<table>
<thead>
<tr>
<th>B</th>
<th>F</th>
<th>D</th>
<th>A</th>
<th>E</th>
<th>C</th>
<th>Row totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>220</td>
<td>98</td>
<td>149</td>
<td>92</td>
<td>282</td>
<td>169</td>
<td>1010</td>
</tr>
<tr>
<td>A</td>
<td>E</td>
<td>B</td>
<td>G</td>
<td>F</td>
<td>D</td>
<td>934</td>
</tr>
<tr>
<td>74</td>
<td>238</td>
<td>158</td>
<td>228</td>
<td>48</td>
<td>188</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>C</td>
<td>F</td>
<td>E</td>
<td>B</td>
<td>A</td>
<td>1034</td>
</tr>
<tr>
<td>118</td>
<td>279</td>
<td>118</td>
<td>278</td>
<td>176</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>B</td>
<td>A</td>
<td>D</td>
<td>C</td>
<td>F</td>
<td>1051</td>
</tr>
<tr>
<td>295</td>
<td>222</td>
<td>54</td>
<td>104</td>
<td>213</td>
<td>163</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>A</td>
<td>B</td>
<td>803</td>
</tr>
<tr>
<td>187</td>
<td>90</td>
<td>242</td>
<td>96</td>
<td>66</td>
<td>122</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>A</td>
<td>C</td>
<td>B</td>
<td>D</td>
<td>E</td>
<td>808</td>
</tr>
<tr>
<td>90</td>
<td>124</td>
<td>195</td>
<td>109</td>
<td>79</td>
<td>211</td>
<td></td>
</tr>
<tr>
<td>Col totals</td>
<td>984</td>
<td>1051</td>
<td>916</td>
<td>907</td>
<td>864</td>
<td>918</td>
</tr>
</tbody>
</table>

(a) Make a complete analysis of this experiment and state which legume you would recommend for planting.

(b) Use the Tukey method (4) to determine if there are significant differences between the three top treatments.

(c) Was the Latin square design useful in this case (see Exercise 9.4.1)?

**Exercise 9.4.4.** Professor John Wishart furnishes us this example:

An experiment on the use of nitrogenous fertilizers on wheat was arranged as a 5 x 5 Latin square, each plot being 1/40 acre in size. The control plot (having no fertilizer) is denoted by 0; S marks the plots which received a single dressing of sulphate of ammonia in March, while SS marks the plots which received the same total dressing, but in 6 monthly dressings from November to April. Plots which received cyanamide in October to an equivalent amount (in nitrogen) as the others are marked C, while D marks plots which received half their dressing as cyanamide and half as dicyanodiamide. The plan is given below, the numbers denoting the yields of the plots in pounds. Conduct the statistical analysis of the data, measuring the significance of the effect of applying a nitrogenous dressing, of whatever kind; also see if you can determine, by a statistical test, whether some forms of dressing are more effective than others. Set out a full summary table with your conclusions.
9.5. The Factorial Design.

9.5.1. The experimenter often wishes to test various types of treatments each with several different representatives. For example, he might test all six treatment combinations of 3 levels of nitrogen fertilizer (f) and 2 varieties (v). These 6 treatments with 5 degrees of freedom can be broken into several pertinent sub-divisions: differences among the 3 fertilizers averaged over the 2 varieties (2 d.f.); differences between the 2 varieties over the 3 fertilizers (1 d.f.); the interactions between fertilizers and varieties which measure the failure of the fertilizers to produce the same results on the 2 varieties (2 d.f.). Let us assume that each of the 6 treatments has been planted on k plots in a randomized complete blocks design. That is each of k blocks contains all 6 treatments. The Latin square design or the completely randomized design are also used as the basic field design for factorials. The analysis of the results differ only in the removal of different block degrees of freedom in computing the error sum of squares.

It is possible to use orthogonal linear forms to compute individual components which represent each of the 5 independent treatment degrees of freedom, each independent of the grand total, G. The experimenter probably wishes to know if the increase in yield with added fertilizer is entirely linear. We can use the (−, 0, +) and (+, −2, +) coefficients to study the linear (F) and quadratic (Fq) effects.

---

See references (1), (8), and (9) for extensive discussions of the analysis and usefulness of various types of factorial designs.
components of the fertilizer effect, each component with one degree of freedom, if the differences between successive levels are the same. For the single degree of freedom for varieties, we shall use the (−, +) coefficients. Then to compute the interaction components, we merely multiply the corresponding main effect coefficients. The 5 orthogonal comparisons plus G are:

<table>
<thead>
<tr>
<th>Orthogonal sets</th>
<th>Original Treatment Totals</th>
<th>Variance of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_\gamma$ = $L_1$</td>
<td>$F_{1V1}$</td>
<td>$F_{1V2}$</td>
</tr>
<tr>
<td>$F_q$ = $L_2$</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>V = $L_3$</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>$F_\kappa$ V = $L_4$</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>$F_q$ V = $L_5$</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>G = $L_6$</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

The totals are computed by adding and subtracting, with proper coefficients, the 6 yields. For example, $F_\gamma$ = (sum of yields for the third level of fertilizer) - (sum of yields for the first level).

We can systematize this procedure by considering that we have p orthogonal linear forms

$$L_h = \sum_{i=1}^{P} a_{hi} T_i; \quad h = 1, 2, \cdots, p,$$

where $\sum_{i=1}^{P} a_{hi} a_{hi} = \begin{cases} 0 & \text{for } h \neq h' \\ 1 & \text{for } h = h' \end{cases}$. The restriction that $\sum_{i=1}^{P} a_{hi}^2 = 1$ requires that each of the coefficients in the (FV) table be divided by the square root of the coefficient of $r \sigma^2$ in the column for the variance of the total; the $T_i$ here refer to the 6 (FV) treatment totals, each with 2 plots. The added (in the table) restriction $\sum_{i=1}^{P} a_{hi} = 0$, for $h = 1, 2, \cdots, p - 1$, was used to make these (p - 1) linear forms independent of the mean equation (G), but is not needed for a general proof.
If we solve for the $T_1$ in terms of the $L_h$, we find that

$$T_1 = \sum_{h' = 1}^{p} a_{h'i} L_{h'}.$$  

This can be derived by matrix algebra, but you can prove it by inserting those values of $T_1$ in any $L_h$ equation:

$$\sum_{i = 1}^{p} a_{hi} T_1 = \sum_{i = 1}^{p} \sum_{h' = 1}^{p} a_{hi} a_{h'i} L_{h'} = L_h,$$

using the summation restrictions on the $a_i$'s. Hence

$$\sum_{i = 1}^{p} T_1^2 = \sum_{i = 1}^{p} \left( \sum_{h' = 1}^{p} a_{h'i} L_{h'} \right)^2 = \sum_{h'} \sum_{i} a_{hi}^2 L_{h'}^2 + \sum_{h' \neq h} \sum_{i} a_{hi} a_{h'i} L_{h'} L_{h''} = \sum_{h' = 1}^{p} L_{h'}^2.$$

Also if $L_p = \frac{G}{\sqrt{p}}$, as in the table above, $\sum_{i = 1}^{p} T_1^2 - \frac{G^2}{p} = \sum_{h' = 1}^{p} L_{h'}^2$.

For those familiar with matrix notation, we indicate

$$\sum T_1^2 = T'T,$$

where $T' = (T_1 T_2 \ldots T_p)$ and $T = \begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_p \end{pmatrix}$. Also $L = AT$, where $A$ is the matrix of the coefficients of the orthogonal transformation. Since the transformation is orthogonal

$$AA' = I \text{ or } A' = A^{-1},$$

where $A'$ is the transpose of $A$ (rows and columns interchanged) and $A^{-1}$ the inverse of $A$. Hence

$$T = A^{-1}L = A'L.$$  

And

$$T'T = (A'L')(A'L) = L'AA'L = L'L.$$  

Hence $\sum T_1^2 = \sum L_{h'}^2$. 

The variance of \( L_i \) is
\[
\sigma^2(L_i) = \sum_i \sum_{i'} a_{hi} a_{hi'}, \quad \sigma^2(T_iT_i') = r \sigma^2,
\]
since \( \sigma^2(T_iT_i') = \begin{cases} 0 & \text{for } i \neq i' \\ r \sigma^2 & \text{for } i = i' \end{cases} \quad \text{and} \quad \sum_i a_{hi}^2 = 1.
\]

\[9.5.2\] Suppose the regression model were set up in the form
\[
Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha \beta)_{ij} + \epsilon_{ijk},
\]
where \( \{\alpha_i\} \) are the fertilizer effects with \( \sum_i \alpha_i = 0 \), \( \{\beta_j\} \) the variety effects with \( \sum_j \beta_j = 0 \), and \( (\alpha \beta)_{ij} \) the additive interaction with
\[
\sum_{i=1}^{3} (\alpha \beta)_{ij} = \sum_{j=1}^{2} (\alpha \beta)_{ij} = 0. \]
The least squares equations for this model would be
\[
\begin{aligned}
F_i &= 2m + 2rf_i = 2r \mu + 2r \alpha_i + \sum_j \sum_k \epsilon_{ijk} \\
V_j &= 3m + 3rv_j = 3r \mu + 3r \beta_j + \sum_i \sum_k \epsilon_{ijk} \\
F_iV_j &= r \left[ m + f_i + v_j + (fv)_{ij} \right] = r \left[ \mu + \alpha_i + \beta_j + (\alpha \beta)_{ij} \right] + \sum_k \epsilon_{ijk} \\
E_k &= 6m + 6b_k = 6(\mu + \beta_k) + \sum_i \sum_j \epsilon_{ijk}.
\end{aligned}
\]
The total treatment sum of squares is
\[
\frac{\sum (F_i - \bar{F})^2}{2r} + \frac{\sum (V_j - \bar{V})^2}{3r} + \frac{\sum \sum (F_iV_j - \frac{F_i}{r} - \frac{V_j}{3} + \frac{G}{b})^2}{r}
= SSF + SSV + SS(FV).
\]

We see that
\[
\begin{aligned}
ESSF &= 2r \sum \alpha_i^2 + 2 \sigma^2 \\
ESSV &= 3r \sum \beta_j^2 + \sigma^2 \\
ESS(FV) &= r \sum \sum (\alpha \beta)_{ij}^2 + 2 \sigma^2.
\end{aligned}
\]

\(2\) We could use two fertilizer constants, one variety constant, and two interaction constants as in 9.5.1. If a completely randomized design were used, \( \beta_k \) would be omitted. And if a Latin square design were used, constants for rows and columns would replace \( \beta_k \).
Also all of these effects are independent.

9.5.3. Hence we have the following analysis of variance

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>EMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>r - 1</td>
<td>SSB</td>
<td>MSB</td>
<td>6 ( \sum \beta_j^2 ) + ( \sigma^2 )</td>
</tr>
<tr>
<td>Fertilizer</td>
<td>2</td>
<td>SSF</td>
<td>MSF</td>
<td>( r \sum \alpha_i^2 ) + ( \sigma^2 )</td>
</tr>
<tr>
<td>Varieties</td>
<td>1</td>
<td>SSV</td>
<td>MSV</td>
<td>3r ( \sum \gamma_j^2 ) + ( \sigma^2 )</td>
</tr>
<tr>
<td>F x V</td>
<td>2</td>
<td>SS(FV)</td>
<td>MS(FV)</td>
<td>( \frac{r \sum (\tau \delta)^2}{2} ) + ( \sigma^2 )</td>
</tr>
<tr>
<td>Error</td>
<td>5(r - 1)</td>
<td>SSE</td>
<td>( s^2 )</td>
<td>( s^2 )</td>
</tr>
</tbody>
</table>

Example 9.5.1.

Consider the following example:

<table>
<thead>
<tr>
<th>F</th>
<th>V</th>
<th>1a</th>
<th>1b</th>
<th>2a</th>
<th>2b</th>
<th>3a</th>
<th>3b</th>
<th>Total</th>
<th>Sub-Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>161</td>
<td>166</td>
<td>113</td>
<td>103</td>
<td>132</td>
<td>180</td>
<td>855</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>192</td>
<td>253</td>
<td>208</td>
<td>171</td>
<td>196</td>
<td>198</td>
<td>1218</td>
<td>2073</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>145</td>
<td>231</td>
<td>131</td>
<td>158</td>
<td>176</td>
<td>216</td>
<td>1057</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>232</td>
<td>231</td>
<td>190</td>
<td>171</td>
<td>242</td>
<td>238</td>
<td>1304</td>
<td>2361</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>172</td>
<td>204</td>
<td>104</td>
<td>135</td>
<td>178</td>
<td>175</td>
<td>968</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>227</td>
<td>214</td>
<td>144</td>
<td>146</td>
<td>186</td>
<td>230</td>
<td>1147</td>
<td>2115</td>
</tr>
<tr>
<td>Total</td>
<td>1129</td>
<td>1299</td>
<td>890</td>
<td>884</td>
<td>1110</td>
<td>1237</td>
<td>6549</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

V1 = 2880
V2 = 3669

The analysis of variance is:

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>5</td>
<td>24,938.92</td>
<td>4,987.78**</td>
</tr>
<tr>
<td>Fertilizers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( F_{X} )</td>
<td>1</td>
<td>73.50</td>
<td>73.50</td>
</tr>
<tr>
<td>( F_{Q} )</td>
<td>1</td>
<td>3,960.50</td>
<td>3,960.50**</td>
</tr>
<tr>
<td>Varieties</td>
<td>1</td>
<td>17,292.25</td>
<td>17,292.25**</td>
</tr>
<tr>
<td>F x V</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( F_{X} \times V )</td>
<td>1</td>
<td>1,410.67</td>
<td>1,410.67</td>
</tr>
<tr>
<td>( F_{Q} \times V )</td>
<td>1</td>
<td>32.00</td>
<td>32.00</td>
</tr>
<tr>
<td>Error</td>
<td>25</td>
<td>9,896.91</td>
<td>395.88</td>
</tr>
</tbody>
</table>

**Significant at 1 percent probability level.
\[ L_1 = F_k = \frac{2115 - 2073}{\sqrt{4}} = 21 \]

\[ L_2 = F_q = \frac{2073 + 2115 - 2(2361)}{\sqrt{12}} = -154.15** \]

\[ L_3 = V = \frac{-2880 + 3669}{\sqrt{6}} = 322.10** \]

\[ L_4 = F_q V = \frac{855 - 1218 - 968 + 1147}{\sqrt{4}} = -92 \]

\[ L_5 = F_q V^2 = \frac{-385 + 1218 + 2(1057 - 1304) - 968 + 1147}{\sqrt{12}} = 13.86 \]

\[ \frac{L_5}{2} = 6 \sqrt{2} = 6(395.88) = 2375.28 \]

\[ \sigma(L_h) = 48.7 \]

The only significant effects are \( F_q \) and \( V \).

Apparently the effect of the fertilizer is definitely quadratic with the maximum yield near the middle rate of application and the yield dropping to almost the same at the high application as at the low application. There is strong evidence of a real difference between the two varieties, and not much evidence of an interaction between fertilizers and varieties. The estimated efficiency of the randomized complete blocks design as compared to a completely randomized design is

\[ \frac{30(395.88) + 24.938.92}{35(395.88)} = 2.65, \]

indicating an expected increase in the error variance of 165 percent if the completely randomized design had been used.

If SSF, SSV and SS(FV) are desired, the following computing procedure is used:

\[ C = (6549)^2/36 = 11,913,722.50. \]

\[ \text{SST} = \left[ \frac{(855)^2 + (1218)^2 + \ldots + (1147)^2}{6} - C \right] = 22,768.92 \]

\[ \text{SSF} = \left[ \frac{(2073)^2 + (2361)^2 + (2115)^2}{12} - C \right] = 4,034.00 \]

\[ \text{SSV} = \left[ \frac{(2880)^2 + (3669)^2}{18} - C \right] = 17,292.25 \]

\[ \text{SS(FV)} = \text{SST} - \text{SSF} - \text{SSV} = 1,442.67 \]
Example 9.5.2. An experiment using 3 varieties of sugar cane (V) and 3 different kinds of nitrogen (N) was conducted in Hawaii in 1942, with 4 replications. The yields in tons of cane per acre were

<table>
<thead>
<tr>
<th>Blocks</th>
<th>V_1 N_1</th>
<th>V_1 N_2</th>
<th>V_1 N_3</th>
<th>V_2 N_1</th>
<th>V_2 N_2</th>
<th>V_2 N_3</th>
<th>V_3 N_1</th>
<th>V_3 N_2</th>
<th>V_3 N_3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>70.5</td>
<td>67.3</td>
<td>79.2</td>
<td>58.6</td>
<td>64.3</td>
<td>64.4</td>
<td>65.8</td>
<td>64.1</td>
<td>56.3</td>
<td>591.2</td>
</tr>
<tr>
<td>2</td>
<td>67.5</td>
<td>75.9</td>
<td>72.8</td>
<td>65.2</td>
<td>48.3</td>
<td>67.3</td>
<td>68.3</td>
<td>64.8</td>
<td>54.7</td>
<td>584.8</td>
</tr>
<tr>
<td>3</td>
<td>63.9</td>
<td>72.2</td>
<td>64.8</td>
<td>70.2</td>
<td>74.0</td>
<td>78.0</td>
<td>72.7</td>
<td>70.9</td>
<td>66.2</td>
<td>632.9</td>
</tr>
<tr>
<td>4</td>
<td>64.2</td>
<td>60.5</td>
<td>86.3</td>
<td>51.8</td>
<td>63.6</td>
<td>72.0</td>
<td>67.6</td>
<td>58.3</td>
<td>54.4</td>
<td>578.7</td>
</tr>
<tr>
<td>Total</td>
<td>266.1</td>
<td>275.9</td>
<td>303.8</td>
<td>245.8</td>
<td>250.2</td>
<td>281.7</td>
<td>274.4</td>
<td>258.1</td>
<td>231.6</td>
<td>2387.6</td>
</tr>
</tbody>
</table>

It is usually best to arrange these results in a 3 x 3 table of means as follows:

<table>
<thead>
<tr>
<th></th>
<th>V_1</th>
<th>V_2</th>
<th>V_3</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>N_1</td>
<td>66.52</td>
<td>61.45</td>
<td>68.60</td>
<td>65.52</td>
</tr>
<tr>
<td>N_2</td>
<td>68.98</td>
<td>62.55</td>
<td>64.52</td>
<td>65.35</td>
</tr>
<tr>
<td>N_3</td>
<td>75.95</td>
<td>70.42</td>
<td>57.90</td>
<td>68.09</td>
</tr>
<tr>
<td>Mean</td>
<td>70.48</td>
<td>64.81</td>
<td>63.67</td>
<td>66.32</td>
</tr>
</tbody>
</table>

\[ G = \frac{(2387.6)^2}{36} = 158350.94 \]

The analysis of variance is:

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>3</td>
<td>200.68</td>
<td></td>
</tr>
<tr>
<td>Varieties</td>
<td>2</td>
<td>319.37</td>
<td>159.69*</td>
</tr>
<tr>
<td>Nitrogen</td>
<td>2</td>
<td>56.54</td>
<td>28.27</td>
</tr>
<tr>
<td>V x N</td>
<td>4</td>
<td>559.79</td>
<td>139.95*</td>
</tr>
<tr>
<td>Error</td>
<td>24</td>
<td>1053.78</td>
<td>43.91</td>
</tr>
</tbody>
</table>

There were significant differences among the average varietal effects, and the average nitrogen effects. However, a significant interaction indicates that the nitrogens have significantly different effects for each variety but in opposite directions. In other words, N_3 is best for V_1 while N_1 is best for V_3. The standard error for the difference between any two of the nine treatments is \[ \sqrt{\frac{2(43.91)}{4}} = 4.68. \] For V_1, N_1 - N_3 (= -9.43) is almost significant, and for V_2, N_1 - N_3 (= 10.70) is significant, at the 5% probability level.
Example 9.5.3. Some data were presented in Exercise 9.2.4 on the gain in weight of rats for four successive generations, both male and female (3). The experimenter not only wants to know if there are differences among the 8 classes but also if these differences can be attributed to sex or generation effects or both. In addition some information regarding the interaction is desired. Because of the unequal numbers of rats for each of the 8 classes, we cannot analyze this experiment in the simple manner presented above for the factorial design with equal numbers (one or more) in the classes. We shall first assume no interaction and test this hypothesis, using the following model:

\[(d') \quad Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk} \quad i = 1, 2; j = 1, 2, 3, 4,\]

where \(\alpha_i\) is the added effect of the \(i^{th}\) sex (\(\alpha_1\) for males and \(\alpha_2\) for females), \(\beta_j\) is the added effect of the \(j^{th}\) generation, and \(\epsilon\) represents the number of rats in the \((i,j)\) class. The error sum of squares (SSE') for \((d')\) is compared with the within classes sum of squares (SSE) for the complete model

\[(d) \quad Y_{ijk} = \mu + \gamma_{ij} + \epsilon_{ijk},\]

where \(\gamma_{ij}\) represents one of the 8 classes.

Using model \((d')\) we have the following least squares equations, where

\[
\hat{Y}_{ij} = m + a_i + b_j,
\]

\[
m: 149 + 55a_1 + 94a_2 + 48b_1 + 40b_2 + 35b_3 + 26b_4 = G = 19,537
\]

\[
a_1: 55m + 55a_1 + 21b_1 + 15b_2 + 12b_3 + 7b_4 = A_1 = 9,203
\]

\[
a_2: 94m + 94a_2 + 27b_1 + 25b_2 + 23b_3 + 19b_4 = A_2 = 10,334
\]

\[
b_1: 48m + 21a_1 + 27a_2 + 48b_1 = B_1 = 6,673
\]

\[
b_2: 40m + 15a_1 + 25a_2 + 40b_2 = B_2 = 5,274
\]

\[
b_3: 35m + 12a_1 + 23a_2 + 35b_3 = B_3 = 4,364
\]

\[
b_4: 26m + 7a_1 + 19a_2 + 26b_4 = B_4 = 3,226
\]

If we want \(m\) to represent the average gain of this experiment, the proper restrictions to put on the equations are:
\[ 55a_1 + 94a_2 = 0; \quad 48b_1 + 40b_2 + 35b_3 + 26b_4 = 0. \]

However, for computational purposes it is easier to first let \( a_2 = 0 \) and \( b_4 = 0 \), change \( a_1 \) to \( a'_1 \), \( b_j \) to \( b'_j \) and \( m \) to \( m' \), and then adjust all results later if the values of \( a_1 \) and \( b_2 \) are wanted. If these latter assumptions are made the \( a_2 \) and \( b_4 \) equations are dropped and the values of \( m' \), \( a'_1 \), \( b'_1 \), \( b'_2 \) and \( b'_3 \) are obtained from the other 5 equations. It is best to solve for these constants by use of the forward solution of the abbreviated Doolittle method (page 142). The computations are

<table>
<thead>
<tr>
<th>( m' )</th>
<th>( a'_1 )</th>
<th>( b'_1 )</th>
<th>( b'_2 )</th>
<th>( b'_3 )</th>
<th>( y )</th>
<th>Check</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m' )</td>
<td>149</td>
<td>55</td>
<td>48</td>
<td>40</td>
<td>35</td>
<td>19,537</td>
</tr>
<tr>
<td>( a'_1 )</td>
<td>55</td>
<td>21</td>
<td>15</td>
<td>12</td>
<td>9,203</td>
<td>9,361</td>
</tr>
<tr>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( b'_1 )</td>
<td>48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( b'_2 )</td>
<td></td>
<td></td>
<td></td>
<td>40</td>
<td>5,274</td>
<td>5,369</td>
</tr>
<tr>
<td>( b'_3 )</td>
<td></td>
<td></td>
<td></td>
<td>35</td>
<td>4,364</td>
<td>4,446</td>
</tr>
</tbody>
</table>

| II     |        |        |        |        |      |        |
| \( A_1 \) | 149    | 55     | 48     | 40     | 35   | 19,537 | 19,864 |
| \( B_1 \) | 1      | 0.3691275 | 0.3221477 | 0.2684564 | 0.2348993 | 131.1208 | 133.3154 |

| III    |        |        |        |        |      |        |
| \( A_2 \) | 34.69799 | 3.281880 | 0.2349000 | -0.9194625 | 1991.356 | 2028.651 |
| \( B_2 \) | 1      | 0.09458415 | 0.006769845 | -0.02649901 | 57.39110 | 58.46595 |

| IV     |        |        |        |        |      |        |
| \( A_3 \) | 32.22650 | -12.90813 | -11.18820 | 190.9497 | 198.9799 |
| \( B_3 \) | 1      | -0.4005440 | -3.471739 | 5.922136 | 6.174418 |

| V      |        |        |        |        |      |        |
| \( A_4 \) | 24.08988 | -13.87112 | 92.12984 | 102.3486 |
| \( B_4 \) | 1      | -0.5758069 | 3.8244041 | 4.248614 |

| A_5     |        |        |        |        |      |        |
| B_5     |        |        |        |        |      |        |

The total sum of squares was \( SY^2 = 2,738,543 \). Since

\[
C = (19,537)^2/149 = (19,537)(131,1208) = 2,562,707.2
\]

\[
Sy^2 = SY^2 - C = 176,835.8
\]

The total reduction due to sex and generations is

\[
SSR = A_{1y}B_{1y} + \cdots + A_{5y}B_{5y} = 115,958.5
\]
Hence

\[ \text{SSE}' = S_y^2 - \text{SSR}' = 60,877.3 \]

The total reduction due to all 8 classes was

\[ \text{SSR} = \frac{(3716)^2}{21} + \cdots + \frac{(2029)^2}{19} - c = 119,141.0 \]

and \( \text{SSE} = S_y^2 - \text{SSR} = 57,694.8 \).

So that the added reduction due to the sex-generation interaction is

\[ \text{SSE}' - \text{SSE} = \text{SSR} - \text{SSR}' = 3,182.5 \]

From the above computations, it is seen that the sum of squares for generations (adjusted for sex) is

\[ \text{SSG} = A_{3y}B_{3y} + A_{4y}B_{4y} + A_{5y}B_{5y} = 1672.4 \]

The sum of squares for sex (unadjusted) is

\[ \frac{(9203)^2}{55} + \frac{(10,334)^2}{94} - c = (1991356)(5739110) = 114,286.1 \]

and the sum of squares for generations (unadjusted) is

\[ \frac{(6673)^2}{48} + \cdots + \frac{(3226)^2}{26} - c = 5,756.4 \]

The following identity holds for these sums of squares:

\[ \text{sex (adjusted for generations)} + \text{generations (unadjusted)} = \text{generations (adjusted for sex)} + \text{sex (unadjusted)} = \text{SSR} \]

Hence the sum of squares for sex (adjusted for generations) is

\[ 115,958.5 - 5,756.4 = 110,202.1 \]

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation (adj. for sex)</td>
<td>3</td>
<td>1,672</td>
<td>557</td>
</tr>
<tr>
<td>Sex (adj. for generation)</td>
<td>1</td>
<td>110,202</td>
<td>110,202**</td>
</tr>
<tr>
<td>Interaction (adj. for sex and generation)</td>
<td>3</td>
<td>3,182</td>
<td>1,061</td>
</tr>
<tr>
<td>Within classes (Error)</td>
<td>141</td>
<td>57,695</td>
<td>409.2</td>
</tr>
</tbody>
</table>
The F-value to test for interaction is

\[ F = \frac{1061}{409.2} = 2.59, \]

with 3 and 141 degrees of freedom, which is not quite significant at the 5% probability level \( F(0.05) = 2.67 \). If we conclude that there is no real interaction, we can use the error mean square (409.2) to test for the adjusted sex and generation effects, showing that there is a highly significant sex difference but no real difference in gains for the four generations. If the interaction is really different, the proper analysis is in question. We might modify model (d') to include \((x^2f)_{ij}\) and then adjust the main effects for the interaction. However, an average main effect has little meaning if interaction is present, because a real interaction indicates that the sex difference is not the same from one generation to another.

For other methods of handling factorial data with unequal numbers in the sub-classes, see the original article (3).
Exercise 9.5.1. An experiment was designed to compare five varieties of cowpeas, at three different spacings 4", 8" and 12" apart in row, with rows 3" apart (10).

Yield of Cowpea Hay (lbs. per 1/100th Morgen Plot)

<table>
<thead>
<tr>
<th>Variety x Spacings</th>
<th>Blocks</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>Total</th>
<th>Subtotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Era 4&quot;</td>
<td></td>
<td>56</td>
<td>45</td>
<td></td>
<td>43</td>
<td>46</td>
<td>190</td>
</tr>
<tr>
<td>8&quot;</td>
<td></td>
<td>60</td>
<td>50</td>
<td></td>
<td>45</td>
<td>48</td>
<td>203</td>
</tr>
<tr>
<td>12&quot;</td>
<td></td>
<td>66</td>
<td>57</td>
<td></td>
<td>50</td>
<td>50</td>
<td>223</td>
</tr>
<tr>
<td>34 C 361 4&quot;</td>
<td></td>
<td>61</td>
<td>58</td>
<td></td>
<td>55</td>
<td>56</td>
<td>230</td>
</tr>
<tr>
<td>8&quot;</td>
<td></td>
<td>60</td>
<td>59</td>
<td></td>
<td>54</td>
<td>54</td>
<td>227</td>
</tr>
<tr>
<td>12&quot;</td>
<td></td>
<td>59</td>
<td>55</td>
<td></td>
<td>51</td>
<td>52</td>
<td>217</td>
</tr>
<tr>
<td>34 C 395 4&quot;</td>
<td></td>
<td>63</td>
<td>53</td>
<td></td>
<td>49</td>
<td>48</td>
<td>213</td>
</tr>
<tr>
<td>8&quot;</td>
<td></td>
<td>65</td>
<td>56</td>
<td></td>
<td>50</td>
<td>50</td>
<td>221</td>
</tr>
<tr>
<td>12&quot;</td>
<td></td>
<td>66</td>
<td>58</td>
<td></td>
<td>52</td>
<td>55</td>
<td>231</td>
</tr>
<tr>
<td>34 C 402 4&quot;</td>
<td></td>
<td>65</td>
<td>61</td>
<td></td>
<td>60</td>
<td>63</td>
<td>249</td>
</tr>
<tr>
<td>8&quot;</td>
<td></td>
<td>60</td>
<td>58</td>
<td></td>
<td>56</td>
<td>60</td>
<td>234</td>
</tr>
<tr>
<td>12&quot;</td>
<td></td>
<td>53</td>
<td>53</td>
<td></td>
<td>48</td>
<td>55</td>
<td>209</td>
</tr>
<tr>
<td>34 C 408 4&quot;</td>
<td></td>
<td>60</td>
<td>61</td>
<td></td>
<td>50</td>
<td>53</td>
<td>224</td>
</tr>
<tr>
<td>8&quot;</td>
<td></td>
<td>62</td>
<td>68</td>
<td></td>
<td>67</td>
<td>60</td>
<td>257</td>
</tr>
<tr>
<td>12&quot;</td>
<td></td>
<td>73</td>
<td>77</td>
<td></td>
<td>77</td>
<td>65</td>
<td>292</td>
</tr>
</tbody>
</table>

Block Totals 929 869 807 815 3420

(a) Set up a table of means.
(b) Set up the analysis of variance.
(c) Derive the linear and quadratic components of the spacing effect. Can you determine the same components of the interaction?
(d) What is the standard error to compare any two of the 15 treatment means?
(e) Discuss the results.

Exercise 9.5.2. A fertility test was made on the growth of grass on Philadelphia Flat soils in the Manti National Forest with three levels of nitrogen (N) and three levels of phosphate (P) with two samples of each treatment.
Grams of Grass

<table>
<thead>
<tr>
<th></th>
<th>N₀</th>
<th>N₁</th>
<th>N₂</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>P₀</td>
<td>18.7</td>
<td>20.8</td>
<td>22.3</td>
<td>61.8</td>
</tr>
<tr>
<td></td>
<td>17.5</td>
<td>20.5</td>
<td>22.9</td>
<td>60.9</td>
</tr>
<tr>
<td></td>
<td>36.2</td>
<td>41.3</td>
<td>45.2</td>
<td>122.7</td>
</tr>
<tr>
<td>P₁</td>
<td>19.2</td>
<td>18.8</td>
<td>24.9</td>
<td>52.9</td>
</tr>
<tr>
<td></td>
<td>21.3</td>
<td>22.5</td>
<td>24.2</td>
<td>67.1</td>
</tr>
<tr>
<td></td>
<td>40.5</td>
<td>42.3</td>
<td>49.1</td>
<td>132.9</td>
</tr>
<tr>
<td>P₂</td>
<td>20.8</td>
<td>22.0</td>
<td>25.6</td>
<td>68.4</td>
</tr>
<tr>
<td></td>
<td>20.5</td>
<td>24.0</td>
<td>27.1</td>
<td>61.6</td>
</tr>
<tr>
<td></td>
<td>41.3</td>
<td>46.0</td>
<td>52.7</td>
<td>140.0</td>
</tr>
<tr>
<td>Total</td>
<td>118.0</td>
<td>129.6</td>
<td>147.0</td>
<td>394.6</td>
</tr>
</tbody>
</table>

Note that this is not a randomized blocks experiment, but is analyzed as a completely randomized design of the type discussed in section 9.2.

(a) Set up the tables of means and make the analysis of variance.

(b) Show that the only important effects are the linear for both N and P.

(c) Discuss the results.

Exercise 9.5.3. Suppose you wish to set up an experiment to test the effectiveness of 2 levels of nitrogen, 2 levels of phosphate and 2 levels of potash on the yield of potatoes and had enough land to plant 80 plots.

(a) Show how you would set up this experiment.

(b) Set up the analysis of variance table.

(c) Indicate what kind of information can be obtained from such an experiment.

(d) If you wanted to know something about the maximum level of the three fertilizers to use, what changes would have to be made in planning another experiment?

(e) How would you take account of the cost of the fertilizers in making your recommendations to the farmer?
Exercise 9.5.4. A 6 x 6 Latin experiment was run at the North Carolina Agricultural Experiment Station to determine the effect of nitrogen and phosphate fertilizers on potato yields. The following treatments were used:

<table>
<thead>
<tr>
<th>Phosphate</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nitrogen</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>High</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
</tbody>
</table>

The field arrangement and yields (lbs. per plot) were:

<table>
<thead>
<tr>
<th></th>
<th>E</th>
<th>B</th>
<th>F</th>
<th>A</th>
<th>C</th>
<th>D</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>633</td>
<td>527</td>
<td>652</td>
<td>390</td>
<td>504</td>
<td>416</td>
<td>3122</td>
<td></td>
</tr>
<tr>
<td>489</td>
<td>475</td>
<td>415</td>
<td>488</td>
<td>571</td>
<td>282</td>
<td>2720</td>
<td></td>
</tr>
<tr>
<td>384</td>
<td>481</td>
<td>483</td>
<td>422</td>
<td>334</td>
<td>646</td>
<td>2750</td>
<td></td>
</tr>
<tr>
<td>620</td>
<td>448</td>
<td>505</td>
<td>439</td>
<td>323</td>
<td>384</td>
<td>2719</td>
<td></td>
</tr>
<tr>
<td>452</td>
<td>432</td>
<td>411</td>
<td>617</td>
<td>594</td>
<td>466</td>
<td>2972</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>505</td>
<td>259</td>
<td>366</td>
<td>326</td>
<td>420</td>
<td>2376</td>
<td></td>
</tr>
<tr>
<td>3078</td>
<td>2868</td>
<td>2725</td>
<td>2722</td>
<td>2652</td>
<td>2614</td>
<td>16659</td>
<td></td>
</tr>
</tbody>
</table>

(a) Make a complete analysis of this experiment, indicating the treatment means and effects and linear and quadratic effects.

(b) What recommendations should be made if it costs $2.00 per plot extra for high instead of low nitrogen and $1.00 per plot extra for each jump in the amount of phosphate and potatoes sold for $0.03 a pound?

(c) Was the Latin square better than a randomized complete blocks design?

Exercise 9.5.5. (a) Use the method presented at the top of page 144 to solve for the values of \( m', a_1', b_1', b_2', \) and \( b_3' \) in Example 9.5.3.

(b) Show that the assumption that \( a_2 = b_4 = 0 \) is equivalent to changing \( \hat{Y} \) to

\[
\hat{Y}_{ij} = m' + a_i' + b_j'; \ i = 1; \ j = 1, 2, 3,
\]

where \( a_i' = a_i - a_2 \) and \( b_j' = b_j - b_4 \), so that \( a_i' = b_j' = 0 \). Use the auxiliary equations \( 55a_1 + 94a_2 = 0 \) and \( 48b_1 + 40b_2 + 35b_3 + 26b_4 = 0 \) to solve for \( m', \{ a_i' \} \) and \( \{ b_j' \} \).
Exercise 9.5.6. C. B. Ratchford of North Carolina State College investigated the
differences in average man work units (MWU) per farm for 114 non-tractor farms in
the Coastal Plains Counties of North Carolina (1949). He studied three factors
which might have influenced the number of MWU per farm: size of farm (small,
medium or large), type of farming (tobacco or general), and three types of rental
arrangements. The total MWU were as follows (number of farms in parentheses):

<table>
<thead>
<tr>
<th>Type of Farming</th>
<th>Size of Farm</th>
<th>Type of Rental Arrangement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tobacco</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Small</td>
<td>7770.7(26)</td>
<td>2337.1(9)</td>
</tr>
<tr>
<td>Medium</td>
<td>2492.5(8)</td>
<td>1239.6(3)</td>
</tr>
<tr>
<td>Large</td>
<td>1654.6(4)</td>
<td>403.3(1)</td>
</tr>
<tr>
<td>Total</td>
<td>11,917.8(38)</td>
<td>4480.0(13)</td>
</tr>
<tr>
<td>General</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>2081.1(9)</td>
<td>1443.7(4)</td>
</tr>
<tr>
<td>Medium</td>
<td>1256.6(4)</td>
<td>1179.0(3)</td>
</tr>
<tr>
<td>Large</td>
<td>680.8(2)</td>
<td>958.7(2)</td>
</tr>
<tr>
<td>Total</td>
<td>4,018.5(15)</td>
<td>3581.4(9)</td>
</tr>
<tr>
<td>Grand Total</td>
<td>15,936.3(53)</td>
<td>8061.4(22)</td>
</tr>
</tbody>
</table>

The total sum of squares within classes was 1,201,641.

(c) Derive the total sum of squares (SSR') due to the main effects of size, type,
and rental arrangement, not adjusted for interaction. Write the variables in
this order in the matrix: rental arrangement, type, and size.

(b) Derive the interaction sum of squares and show that there was no significant
interaction.

(c) Test the effect of size adjusted for rental arrangement and type of farming.

(d) How would you go about testing the other two adjusted main effects?

Exercise 9.5.7. Read an article by Andorson and Manning (11) for a more complete
discussion of matrix methods with unequal frequencies.
9.6. Balanced Incomplete Blocks Designs

9.6.1. Suppose we have \( p \) treatments each applied to \( r \) plots but with only \( k \) treatments included in each block \( (k < p) \). Assume that the total number of blocks is \( q \) \( (q > r) \) so that the total number of plots, \( N = kq = rp \). The \( i \)th treatment appears in the \( j \)th block \( n_{ij} \) times \( (n_{ij} = 0 \) or \( 1) \). Hence \( \sum_j n_{ij} = r \) and \( \sum_i n_{ij} = k \). This design is called an incomplete blocks design. If a further restriction is made on the design that every treatment appears with every other treatment in the same block an equal number of times, say \( \lambda \), the incomplete blocks design is said to be balanced. The model for this design is

\[
Y_{ij} = n_{ij}(\mu + \tau_i + \beta_j + \epsilon_{ij}), \quad i = 1, 2, \ldots, p, \quad j = 1, 2, \ldots, q,
\]

where \( \sum_i \tau_i = \sum_j \beta_j = 0 \). It might be noted that there are two important situations for which these incomplete blocks design are used:

(a) The number of possible plots per block is less than \( p \). Examples of this are:

(i) Nutrition studies on animals where the block is a litter, the individual animal is the plot and the smallest litter size is smaller than the number of rations to be studied.

(ii) Chemical experiments where the block is a day, an individual analysis is the plot, and the number of possible analyses in a day is less than the number of treatments.

(iii) Tasting experiments where the block is a single trial by an individual taster, the score on a given product is the plot yield, and the number of different products which a taster can differentiate on a single trial is less than the number of products being tasted.

(iv) Education or psychology tests in which the block is the trials by a single child on a given day, the plot is a single test on this day and the number of tests being considered is greater than the number the child can take on a given day without tiring.
(b) The number of treatments is so large that enough homogeneous plots cannot be
found for a complete block (a block containing enough plots for all treatments)\(^6\).

(i) If we have 81 varieties of corn to test, it is often impossible to find
81 more or less homogeneous plots to form a complete block.

(ii) For greenhouse experiments, temperature and light conditions vary so much
that it is desirable to form blocks of rather small size, the size often
being smaller than needed to include all of the pots for a complete block.

We will not include many examples of these designs in this book, because of a
limitation of space. Students interested in more examples should consult references
(1), (12) and (13).

9.6.2. The least squares equations for the balanced incomplete blocks model are:

\[
\begin{align*}
\mathbf{b}_j & : \mathbf{k}m + \sum_i n_{ij}t_i + \mathbf{k}b_j = \mathbf{B}_j \\
\mathbf{t}_i & : \mathbf{r}m + \mathbf{r}t_i + \sum_j n_{ij}b_j = \mathbf{T}_i \\
A_i &= \mathbf{T}_i - \frac{1}{k} \sum_j n_{ij}B_j = \mathbf{r}t_i - \frac{1}{k} \sum_{j=1}^{q} \sum_{\ell=1}^{p} n_{ij}\lambda_{j\ell}\mathbf{t}_\ell \\
&= (r - \frac{r}{k})t_i - \frac{\lambda}{k} \sum_{\ell \neq i} \mathbf{t}_\ell = \frac{r(k - 1) + \lambda}{k} \mathbf{t}_i,
\end{align*}
\]

since \(\sum_{i=1}^{p} t_i = 0\) and \(\sum_{j=1}^{q} b_j = 0\) (so that \(\sum_{\ell \neq i} \mathbf{t}_\ell = -\mathbf{t}_i\)).

If we set \(\frac{r(k - 1) + \lambda}{k} = rE_F\),

\[
t_i = A_i/rE_F.
\]

\(E_F\) is called the efficiency factor for the incomplete blocks design, since the
number of effective replications is \(rE_F\) instead of \(r(\frac{E_F}{r} < 1)\). We see that the
number of blocks is given by

\[
q = \frac{\lambda_p c_2}{k c_2} = \frac{\lambda_p(p - 1)/k(k - 1)}
\]

\(^6\) A complete block is usually called a replication by experimental statisticians.
Hence \[ p(p - 1) = qk(k - 1) = rp(k - 1), \] and

\[ E_F = \frac{r(k - 1) + \lambda}{rk} = \frac{\lambda p}{rk} \frac{p(k - 1)}{k(p - 1)}. \]

We might say that a necessary condition for "balance" is that \( \lambda p(p - 1) = qk(k - 1) = rp(k - 1) \). However, it may not be possible to construct a balanced design even though this condition is met.

### 9.6.3

The adjusted mean yield for the \( i \)th treatment is

\[ t_i' = t_i + m = A_i/rE_F + G/rp = \left[ k(p - 1)A_i + (k - 1)G \right]/rp(k - 1). \]

In section 8.5.2 (page 144), it was indicated that the reduction due to any variate, \( X_i \), adjusted for all previous variates is given by \( A_{iY}B_{iY} \), where \( A_{iY} = S_Y(x_i) \)

adjusting for all previous independent variates in the matrix) and \( B_{iY} \) was the regression coefficient for \( X_i \). In the balanced incomplete blocks design, the regression coefficient, \( t_i = A_i/rE_F \) and \( A_i \) is the sum of cross-products, adjusted for block effects. Hence the treatment sum of squares adjusted for block effects is

\[ SST(\text{adj}) = \sum A_i^2/rE_F. \]

A more formal proof would be the following:

The total reduction in sum of squares due to blocks and treatments is

\[ SSR = \sum b_iB_j + \sum t_iT_i. \]

**SSR** (omitting treatments) = \( \sum (B_j - \bar{B})^2/k = \sum B_j(B_j - \bar{B})/k \)

**SST** (adj.) = **SSR** - **SSR**

\[
= \sum b_j - \left( \frac{\bar{B} + \bar{B}}{k} \right)B_j + \sum A_i(\bar{A}_i + \frac{1}{k} \sum n_{ij}B_j)/rE_F
\]

\[
= \sum A_i^2/rE_F + \sum B_j \left[ b_j + \frac{1}{k} \sum n_{ij}t_i - \frac{B_i - \bar{B}}{k} \right],
\]

where \( \bar{B} = G/q \). From the \( G \) and \( B_j \) equations, we see that

\[ b_j + \frac{1}{k} \sum n_{ij}t_i = \frac{B_i - \bar{B}}{k}. \]
Hence \( \text{SST(adj.)} = \sum_i \frac{A_i^2}{rE_f} \). Also \( \text{SSE} = \text{Sy}^2 - \text{SSR} \) with \( (N - p - q + 1) \) degrees of freedom.

It might be noted that

\[
B_j - \bar{B} = k \beta_j + \sum_i n_{ij} T_i + \left[ \sum_i n_{i} \xi_{ij} \right]
\]

\[
A_i = rE_f T_i + \left[ \sum_j n_{ij} \xi_{ij} \right]
\]

\[
G = \frac{r \mu}{N} + \sum_i \sum_j n_{ij} \xi_{ij}
\]

It can be shown that

\[
E(B_j - \bar{B})^2 = (k \beta_j + \sum_i n_{ij} T_i)^2 + k(q - 1) \sigma^2 / q
\]

\[
E(A_i^2) = (rE_f)^2 T_i^2 + r(k - 1) \sigma^2 / k
\]

\[
E(G^2) = (r \mu)^2 + r \mu \sigma^2
\]

\[
\sigma^2 (A_i A_j) = - \lambda \sigma^2 / k \quad (i \neq j)
\]

\[
\sigma^2 (B_j - \bar{B})A_i = \sigma^2 (B_j - \bar{B})G = \sigma^2 (A_i G) = 0.
\]

The analysis of variance is

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>E.M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>( q - 1 )</td>
<td>( \sum (B_j - \bar{B})^2 / k(q - 1) )</td>
<td>( \sigma^2 + \theta { T_i, \beta_j } )</td>
</tr>
<tr>
<td>Treatments (adj.)</td>
<td>( p - 1 )</td>
<td>( \sum A_i^2 / rE_f (p - 1) )</td>
<td>( \sigma^2 + rE_f \sigma^2 / t )</td>
</tr>
<tr>
<td>Error</td>
<td>( N - p - q + 1 )</td>
<td>( s^2 )</td>
<td>( \sigma^2 )</td>
</tr>
</tbody>
</table>

where \( \theta \) is a function of all the \( T_i \)'s and \( \beta_j \)'s and \( \sigma^2 = \sum T_i^2 / (p - 1) \).

2.6.4. It is easy to see that the one-tailed F-test is appropriate to test

\[
H_0 \quad \left\{ \sum T_i = 0 \right\}
\]

with \( s^2 \) estimated by the error sum of squares divided by \( (N - p - q + 1) \).
\[ \sigma^2(t_1) = (k - 1) \sigma^2/rkE_r^2 \]
\[ \sigma^2(t_1') = \sigma^2(t_1) + \sigma^2/rp \]
\[ \sigma^2(t_1 - t_2) = \sigma^2(t_1' - t_2') = \left[ 2r(k - 1) + 2 \lambda \right] \sigma^2/k(rE_r)^2 = 2 \sigma^2/rE_r. \]

**Example 9.6.1.** A special type of a balanced incomplete blocks design is one with \( p = k^2 \) treatments, a balanced lattice design. This design can be set up in \( r \) complete replications of \( k \) blocks each (with \( k \) treatments per block); hence,
\[ q = rk, \quad \lambda = r(k - 1)/p - 1 = r/(k + 1), \quad \text{and} \quad E_r = k/(k + 1). \]
A balanced lattice design with \( p = 9(k = 3) \) and \( r = 4 \) was used to test 9 rations fed to rats. Hence \( \lambda = 1 \) and \( E_r = 3/4 \). The gains for this experiment were (the ration numbers are in parentheses):

<table>
<thead>
<tr>
<th>j</th>
<th>Replication I</th>
<th>Replication II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bj</td>
<td>Bj</td>
</tr>
<tr>
<td>1</td>
<td>(1) 20 (4) 15 (7) 11 46</td>
<td>4 (7) 8 (8) 12 (9) 16 36</td>
</tr>
<tr>
<td>2</td>
<td>(3) 8 (6) 18 (9) 26 52</td>
<td>5 (1) 20 (2) 2 (3) 2 24</td>
</tr>
<tr>
<td>3</td>
<td>(2) 18 (5) 16 (8) 2 36</td>
<td>6 (4) 20 (5) 6 (6) 2 28</td>
</tr>
<tr>
<td>Total</td>
<td>134</td>
<td>Total</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>j</th>
<th>Replication III</th>
<th>Replication IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bj</td>
<td>Bj</td>
</tr>
<tr>
<td>7</td>
<td>(1) 13 (9) 19 (5) 14 46</td>
<td>9 (5) 19 (7) 23 (3) 6 48</td>
</tr>
<tr>
<td>8</td>
<td>(8) 14 (4) 34 (3) 2 50</td>
<td>10 (1) 22 (6) 12 (8) 2 36</td>
</tr>
<tr>
<td>9</td>
<td>(6) 14 (2) 20 (7) 14 48</td>
<td>11 (9) 27 (2) 7 (4) 20 54</td>
</tr>
<tr>
<td>Total</td>
<td>144</td>
<td>Total</td>
</tr>
</tbody>
</table>

The computations are as follows:

\[
\text{Ration} \quad \text{Total Gain} = T_1 = \sum_{j=1}^{12} n_{ij} B_j = T_{b1} \quad 3a_1 = 3T_1 - T_{b1} \quad t'_1 = (12a_1 + G)/36
\]

<table>
<thead>
<tr>
<th>Ration</th>
<th>Total Gain</th>
<th>( \sum_{j=1}^{12} n_{ij} B_j = T_{b1} )</th>
<th>3a_1</th>
<th>t'_1</th>
<th>(12a_1 + G)/36</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>75</td>
<td>152</td>
<td>73</td>
<td>22.11</td>
<td>504 = G</td>
</tr>
<tr>
<td>2</td>
<td>47</td>
<td>162</td>
<td>-21</td>
<td>11.67</td>
<td>1512 = 3G</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>174</td>
<td>-120</td>
<td>0.67</td>
<td>1512 = 3G</td>
</tr>
<tr>
<td>4</td>
<td>89</td>
<td>178</td>
<td>-89</td>
<td>23.89</td>
<td>1512 = 3G</td>
</tr>
<tr>
<td>5</td>
<td>55</td>
<td>158</td>
<td>7</td>
<td>14.78</td>
<td>1512 = 3G</td>
</tr>
<tr>
<td>6</td>
<td>46</td>
<td>164</td>
<td>-26</td>
<td>11.11</td>
<td>1512 = 3G</td>
</tr>
<tr>
<td>7</td>
<td>56</td>
<td>178</td>
<td>-10</td>
<td>12.39</td>
<td>1512 = 3G</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
<td>158</td>
<td>-68</td>
<td>6.44</td>
<td>1512 = 3G</td>
</tr>
<tr>
<td>9</td>
<td>88</td>
<td>188</td>
<td>76</td>
<td>22.44</td>
<td>1512 = 3G</td>
</tr>
<tr>
<td>Total</td>
<td>504 = G</td>
<td>1512 = 3G</td>
<td>0</td>
<td>126.00</td>
<td></td>
</tr>
</tbody>
</table>

A more complicated lattice with \( p = k^3 \) treatments is called a cubic lattice.
In this experiment, a constant for replications can be inserted in the model and a sum of squares for replications removed from the block sum of squares. The total sum of squares for blocks is

$$SSB = \frac{(46)^2 + (52)^2 + \cdots + (36)^2 + (54)^2}{3} - \frac{(504)^2}{36}$$

$$= 7402.67 - 7056 = 346.67$$

The sum of squares for replications is

$$\frac{(134)^2 + \cdots + (138)^2}{9} - 7056 = 219.56$$

Hence the sum of squares for blocks in replications is

$$346.67 - 219.56 = 127.11$$

The sum of squares for treatments (adjusted for blocks) is

$$SST_{(adj_1)} = \frac{(73)^2 + (-21)^2 + \cdots + (76)^2}{(9)(3)} = 1456.15$$

The error sum of squares is

$$S_y^2 - SSB - SST_{(adj_1)} = 9316 - 7056 - 346.67 - 1456.15 = 457.18$$

Hence the analysis of variance table is

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>E.M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replications</td>
<td>3</td>
<td>219.56</td>
<td>73.19</td>
<td></td>
</tr>
<tr>
<td>Blocks (in reps.)</td>
<td>8</td>
<td>127.11</td>
<td>15.89</td>
<td></td>
</tr>
<tr>
<td>Treatments (adj.)</td>
<td>8</td>
<td>1456.15</td>
<td>182.02</td>
<td>(\sigma^2 + 3 \sigma^2_t)</td>
</tr>
<tr>
<td>Error</td>
<td>16</td>
<td>457.18</td>
<td>28.57</td>
<td>(\sigma^2)</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>2260</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To test for treatment differences, \(F = 182.02/28.52 = 6.37\) with 8 and 16 degrees of freedom, for which \(P < .001\). The standard error of the difference between any two adjusted treatment means \((t_1^t - t_2^t)\) is

$$\sqrt{\frac{2(28.57)}{3}} = 4.36$$
We can compute the efficiency of this design relative to a randomized complete blocks design with 4 replications (complete blocks). First we compute SST (unadjusted) and then the estimated randomized blocks error as the difference $\bar{S}^2 = \text{SST} - \text{SS (replications)}$.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replications</td>
<td>3</td>
<td>219.56</td>
<td></td>
</tr>
<tr>
<td>Treatments</td>
<td>8</td>
<td>1194.00</td>
<td></td>
</tr>
<tr>
<td>Randomized blocks error</td>
<td>24</td>
<td>846.44</td>
<td>35.27</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>2260.00</td>
<td></td>
</tr>
</tbody>
</table>

The effective incomplete blocks error is $\frac{28.57}{38.09} = 0.749$. Hence the efficiency of the incomplete blocks design is only $\frac{35.27}{38.09} = 0.926$. In addition there is another loss due to fewer error degrees of freedom (see (1), page 28).

Exercise 9.6.1. A simple example of a balanced incomplete blocks experiment is the following balanced lattice with four treatments, six blocks and two treatments per block.

<table>
<thead>
<tr>
<th>Treatments</th>
<th>Ia</th>
<th>Ib</th>
<th>IIa</th>
<th>IIb</th>
<th>IIIa</th>
<th>IIIb</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>30</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td>85</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td></td>
<td>10</td>
<td>20</td>
<td></td>
<td></td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>30</td>
<td>25</td>
<td></td>
<td>40</td>
<td></td>
<td></td>
<td>95</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td></td>
<td>10</td>
<td>15</td>
<td></td>
<td></td>
<td>35</td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>40</td>
<td>55</td>
<td>20</td>
<td>40</td>
<td>60</td>
<td>260</td>
</tr>
</tbody>
</table>

(a) Set up the analysis of variance and adjusted treatment means for this experiment, and test for treatment differences.

(b) Estimate the standard error of the difference between any two adjusted treatment means.

(c) Show how the replication constants fit into the model and determine the expected value of the replication mean square in the analysis of variance.

(d) Compute an estimate of the efficiency of this design relative to a randomized complete blocks design.
Exercise 9.6.2. Dr. Pauline Paul of Iowa State College conducted an experiment to compare the effects of cold storage on the tenderness and flavor of beef roasts (1). Six periods of storage (0, 1, 2, 4, 9, and 18 days) were tested (p = 6). Since the same cut of meat on each side of an animal was expected to be similar but different cuts on the same side dissimilar, it was decided to use a balanced incomplete blocks design with k = 2, \( \lambda = 1, \) q = 15, and r = 5. In this case it was also possible to arrange the cuts in complete replication of 3 cuts from each side. The scores for tenderness of beef were (periods of storage in parentheses):

<table>
<thead>
<tr>
<th>Replication I</th>
<th>Replication II</th>
<th>Replication III</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0) 7 (1) 17</td>
<td>(0) 17 (2) 27</td>
<td>(0) 10 (4) 25</td>
</tr>
<tr>
<td>(2) 26 (4) 25</td>
<td>(1) 23 (9) 27</td>
<td>(1) 26 (18) 37</td>
</tr>
<tr>
<td>(9) 33 (18) 29</td>
<td>(4) 29 (18) 30</td>
<td>(2) 24 (9) 26</td>
</tr>
<tr>
<td></td>
<td>137</td>
<td>153</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Replication IV</th>
<th>Replication V</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0) 25 (9) 40</td>
<td>(0) 11 (18) 27</td>
</tr>
<tr>
<td>(1) 25 (4) 34</td>
<td>(1) 24 (2) 21</td>
</tr>
<tr>
<td>(2) 34 (18) 32</td>
<td>(4) 26 (9) 32</td>
</tr>
<tr>
<td>190</td>
<td>141</td>
</tr>
</tbody>
</table>

The blocks run across the page.

(a) Make a complete analysis of these data, showing the adjusted treatment means, the analysis of variance, the standard error of treatment differences, and general conclusions.

(b) Show how to determine the number of paired cuts of beef (blocks) needed for a balanced experiment in this case.

(c) Suppose it had been possible to pair the cuts into groups of four like ones instead of two. What design would you then set up?

(d) Is there any method of determining the trend of tenderness in terms of storage time?

(e) Was there any appreciable gain in using the incomplete blocks design for this problem?

Exercise 9.6.3. Moore and Bliss (14) set up a balanced incomplete blocks design to compare the toxicity of each of 7 chemicals on *Aphis rumicis*. The measure of the toxicity was the logarithm of the dose (+3.806) required for a 95 percent kill.
Since only three chemicals could be tested on a given day, \( k = 3 \). 7 days were required to make a balanced design. The toxicities were as follows:

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Day 3</th>
<th>Day 4</th>
<th>Day 5</th>
<th>Day 6</th>
<th>Day 7</th>
<th>( T_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>465</td>
<td>602</td>
<td>443</td>
<td>652</td>
<td>536</td>
<td></td>
<td></td>
<td>1.910</td>
</tr>
<tr>
<td>B</td>
<td>343</td>
<td></td>
<td>873</td>
<td>875</td>
<td>1.142</td>
<td></td>
<td></td>
<td>1.531</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>396</td>
<td>325</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.890</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td>609</td>
<td></td>
<td></td>
<td></td>
<td>1.330</td>
</tr>
<tr>
<td>E</td>
<td></td>
<td>634</td>
<td></td>
<td>409</td>
<td>417</td>
<td></td>
<td></td>
<td>1.600</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td>987</td>
<td>989</td>
<td>931</td>
<td></td>
<td></td>
<td>2.907</td>
</tr>
<tr>
<td>G</td>
<td></td>
<td>330</td>
<td>426</td>
<td>309</td>
<td></td>
<td></td>
<td></td>
<td>1.065</td>
</tr>
</tbody>
</table>

| BJ       | 2.04  | 2.109 | 1.530 | 1.856 | 2.783 | 1.254 | 1.957 | 12.693  |

(a) Can you separate out replications in the analysis of those data?
(b) Show how to determine the number of days required for a balanced experiment.
(c) Make a complete analysis of those data.
(d) C and F were basically the same chemical compound. Were they different from one another? How did they compare with the standard treatment, A?
(e) How did A compare with all the others, excepting C and F?
2.7. Non-Balanced Incomplete Blocks Designs.

2.7.1. If every treatment in 2.6 does not appear the same number of times in the same block with every other treatment, the design is not balanced. Suppose the $i$th treatment appears $\lambda_{i_1}^L$ times in the same block with the $L$th treatment. Then

$$A_i = \left[ r(k - 1)t_i - \sum_{L \neq i} \lambda_{iL}^L t_L \right] / k \quad (i, L = 1, 2, \cdots, p),$$

where $\sum A_i = 0$. The values of the $t_i$ can only be solved by the simultaneous solution of the $p A_i$ equations, with the linear restraint, $\sum t_i = 0$. One method of solving these $p$ equations is to set $t_i = t_i'' + t_p$; so that $t_p'' = 0$, and $t_p = - \sum t_i''/p$. Hence

$$A_i = \left[ r(k - 1)t_i'' - \sum_{L \neq i} \lambda_{iL}^L t_L'' \right] / k \quad (i = 1, 2, \cdots, p - 1)$$

since $\sum_{L \neq i} \lambda_{iL}^L = r(k - 1)$. We see that

$$t_i'' = \sum_{L = 1}^{p - 1} c_{iL}^L A_L^L,$$

where the $c_{iL}^L$ are the elements of the inverse matrix of the coefficients of $t_i''$ in the $(p - 1) A_i$ equations.

In this case

$$\text{SST(adj.)} = \sum_{i = 1}^{p - 1} t_i'' A_i = \sum_{i = 1}^{p - 1} (t_i - t_p) A_i = \sum_{i = 1}^{p - 1} t_i A_i.$$ 

It can be shown that

$$E \left[ \text{MST(adj.)} \right] = \sigma^2 + \theta \left\{ \frac{\sum u_i}{i = 1} \right\} / (p - 1),$$

where $\theta$ is 0 when $\left\{ u_i = 0 \right\}$. Also

$$S^2(t_i'') = c_{ii}^L \sigma^2 \quad \text{and} \quad S^2(t_i'' - t_L'') = (c_{ii}^L - 2c_{iL}^L + c_{L}^L) + \sigma^2,$$

where $\sigma^2 = \sigma^2 = \text{SSE}/(N - p - q + 1)$.

2.7.2. The simple lattice designs. One special type of the non-balanced incomplete blocks design is a lattice design with fewer replications than are needed for balance. A lattice design with two replications is called a simple lattice, one
with three replications a triple lattice, etc. It was shown in Example 9.6.1 that a lattice is balanced if the number of replications, \( r \leq k + 1 \). Hence when \( r < k + 1 \), the lattice is not balanced. The analysis of a non-balanced lattice is much simpler than the general analysis presented in 9.7.1., because of a special method of allocating the treatments to the blocks. The \( k^2 \) treatments are assigned at random one of the following treatment numbers.

\[
\begin{bmatrix}
11 & 12 & \cdots & 1k \\
21 & 22 & \cdots & 2k \\
\vdots & \vdots & & \vdots \\
k1 & k2 & \cdots & kk \\
\end{bmatrix}
\]

If a simple lattice is used \((r = 2)\), the row combinations \((11, 12, \cdots, 1k; 21, 22, \cdots, 2k; \text{etc.)}\) are assigned to separate blocks in one replication, \(X\), and the column combinations \((11, 21, \cdots, k1; 12, 22, \cdots, k2; \text{etc.)}\) assigned to separate blocks in the second replication, \(Y\). If a triple lattice is used, an additional replication, \(Z\), is taken from the diagonals \((11, 22, \cdots, kk; 21, 32, \cdots, 1k; \text{etc.)}\). For a complete discussion of this design, see references (12) and (15).

Let us consider the analysis of a simple lattice experiment \((r = 2)\). We shall designate the treatment effects as \( t_{ij} \) \((i, j = 1, 2, \cdots, k)\) and similarly for treatment totals \((T_{ij})\) and adjusted treatment totals \((A_{ij})\). Let

\[
a_i = \sum_j t_{ij} \quad \text{and} \quad d_j = \sum_i t_{ij}.
\]

The yield of the \((i, j)\) treatment in the \(X\) replication is designated as \(X_{ij}\) and in the \(Y\) replication as \(Y_{ij}\). Therefore \(T_{ij} = X_{ij} + Y_{ij}\). And finally we shall use the notation

\[
\sum_i X_{ij} = X_{.j} \quad ; \quad \sum_j X_{ij} = X_{.i} \quad ; \quad \sum_i \sum_j X_{ij} = X_{..}
\]

and similarly for \(Y\). Hence

\[
A_{ij} = T_{ij} - \frac{1}{k}(X_{.i} + Y_{.j}).
\]
Also we note that $\lambda$ will be 1 for those treatments with the same row or column subscript and 0 elsewhere. Hence

$$A_{ij} = \left[ 2(k-1)t_{ij} - \sum_{j' \neq j} t_{ij'} - \sum_{i' \neq i} t_{i'j} \right] / k$$

$$= \left[ 2t_{ij} - (a_i + d_j) / k \right].$$

$$\sum_j A_{ij} = \left[ Y_{ij} - Y_{..} / k \right] = a_i; \quad \sum_i A_{ij} = \left[ X_{..} - X_{ij} / k \right] = d_j,$$

since $\sum a_i = \sum d_j = \sum \sum t_{ij} = 0$.

We have shown that the sum of squares for treatments (adjusted for blocks) is

$$\text{SST(adj.)} = \sum \sum A_{ij} t_{ij} = \frac{1}{2} \sum \sum A_{ij} \left[ A_{ij} + (a_i + d_j) / k \right]$$

$$= \frac{1}{2} \sum \sum \left[ T_{ij} - (X_{..} + X_{ij}) / k \right] \left[ T_{ij} - (X_{..} + Y_{..} - X_{ij} - Y_{ij}) / k \right].$$

By expanding those products term by term, we have

$$\text{SST(adj.)} = \frac{1}{2} \left\{ \sum \sum T_{ij}^2 - (X_{..} - Y_{..})^2 / k^2 \right.$$

$$\left. - \left[ \sum_{i} T_{i..}^2 + \sum_{j} T_{..j}^2 \right] / k + 2 \left[ \sum_{i} Y_{i..}^2 + \sum_{j} x_{ij}^2 \right] / k \right\},$$

where $T_{i..} = X_{i..} + Y_{i..}$ and $T_{..j} = X_{..j} + Y_{..j}$.

The error sum of squares is obtained by computing the unadjusted block sum of squares, SSB, and then subtracting to obtain

$$\text{SSE} = S y^2 - \text{SSB} - \text{SST(adj.)}$$

The analysis of variance is

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>$2k - 1$</td>
<td>SSB</td>
<td>MST(adj.)</td>
</tr>
<tr>
<td>Treatments (adj.)</td>
<td>$k^2 - 1$</td>
<td>SST(adj.)</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>$(k - 1)^2$</td>
<td>SSB</td>
<td>s^2</td>
</tr>
</tbody>
</table>

The blocks sum of squares can be divided into two parts, for replications ($1 \text{ d.f.}$) and blocks in replications ($2(k - 1) \text{ d.f.}$). The replication part is independent of treatment effects and equals $(X_{..} - Y_{..})^2 / 2k^2$. The blocks-in-replication part is a mixture of block and treatment effects as well as error and is simply

$$\text{SSB} - (X_{..} - Y_{..})^2 / 2k^2.$$
In this case it is not possible to give a single figure for the variance of the difference between two adjusted means, since some treatments will appear together in the same block and others never appear together in the same block. It seems reasonable that the variance of the difference will be lower for those treatments appearing together in the same block. Let us first consider two adjusted treatment effects $t'_{1j}$ and $t'_{2j}$, which appear in the same block.

$$t'_{1j} = \frac{1}{k} \left[ X_{1j} + Y_{1j} - (X_{1*} - Y_{1*} - X_{*j} + Y_{*j})/k \right]$$

$$t'_{1j} - t'_{2j} = \frac{1}{k} \left[ X_{1j} + Y_{1j} - X_{2j} - Y_{2j} - (X_{1*} - X_{2*} - Y_{1*} + Y_{2*})/k \right]$$

$$= \frac{1}{k} \left[ (\frac{k-1}{k}) (X_{1j} - X_{2j}) - (\frac{1}{k}) \sum_{j' \neq j} (X_{1j'} - X_{2j'}) \right]$$

$$+ (\frac{k+1}{k}) (Y_{1j} - Y_{2j}) + (\frac{1}{k}) \sum_{j' \neq j} (Y_{1j'} - Y_{2j'})$$

$$\sigma^2(t'_{1j} - t'_{2j}) = \frac{2}{4k^2} \left[ (k-1)^2 + (k-1) + (k+1)^2 + (k-1) \right] \sigma^2$$

$$= (k+1) \sigma^2/k$$

Now consider two treatments which do not appear together in the same block, such as $t_{11}$ and $t_{22}$. In this case the $X_{*j}$ and $Y_{*j}$ do not cancel out, as above, and we have

$$\sigma^2(t'_{11} - t'_{22}) = (k+2) \sigma^2/k.$$ 

We note that there are $k^2(k^2 - 1)/2$ possible treatment comparisons, of which $k^2(k-1)$ are between treatments in the same block and $k^2(k-1)^2$ between treatments not in the same block. Hence we might use as an average variance of the difference between adjusted treatment effects

$$\frac{k^2(k-1)(k+1)/k + k^2(k-1)^2(k+2)/2k}{k^2(k^2 - 1)/2} \sigma^2 = (k+3) \sigma^2/(k+1)$$

The factor $(k+1)/(k+3)$ is the efficiency factor, $E_f$, for this design.
To illustrate the computing techniques for the simple lattice, we shall use the first two replications of the \((3 \times 3)\) experiment considered in Example 9.6.1. We shall designate the treatments as 11, 12, 13, \(\cdots\), 33 instead of 1, 2, \(\cdots\), 9, with \(X\) for Replication II and \(Y\) for replication I. The treatment and \(X\) and \(Y\) totals are (treatment numbers in parentheses):

\[
\begin{array}{cccc}
(11) & 40 & (12) & 20 \\
(21) & 35 & (22) & 22 \\
(31) & 19 & (32) & 14 \\
\hline
T_{i.} & & & \\
\end{array}
\]

\[
\begin{array}{cccc}
(13) & 10 & (14) & 70 \\
(23) & 20 & (24) & 77 \\
(33) & 42 & (34) & 75 \\
\hline
T_{.j} & 94 & 56 & 72 222 \\
\end{array}
\]

\[
\begin{array}{cccc}
1 & 48 & 46 & 24 & 46 \\
2 & 20 & 49 & 28 & 36 \\
3 & 20 & 39 & 36 & 52 \\
\hline
\text{Total} & 88 & 134 & 88 & 134 \\
\end{array}
\]

\[
S_{Y}^2 = 3706 - \frac{(222)^2}{18} = 968
\]

\[
SST(\text{adj.}) = \frac{1}{2} \left\{ (40)^2 + \cdots + (42)^2 - \frac{(88 - 134)^2}{9} - \frac{[70^2 + 77^2 + \cdots + 72^2]}{3} \right\}
\]

\[
\begin{aligned}
&+ \frac{2}{3} \left[ (48)^2 + (20)^2 + \cdots + (39)^2 \right] \\
&= \frac{1}{2} \left[ 6530 - 235.11 - 11,203.33 + 6,094.67 \right] = 593.11
\end{aligned}
\]

\[
SSB = \frac{(24)^2 + (28)^2 + \cdots + (52)^2}{3} - \frac{(222)^2}{18} = 186.
\]

The analysis of variance is

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replications</td>
<td>1</td>
<td>118</td>
<td></td>
</tr>
<tr>
<td>Blocks (in replications)</td>
<td>4</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>Treatments (adj.)</td>
<td>8</td>
<td>593</td>
<td>74</td>
</tr>
<tr>
<td>Error</td>
<td>4</td>
<td>189</td>
<td>47.2</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>968</td>
<td></td>
</tr>
<tr>
<td>Treatments (unadjusted)</td>
<td>8</td>
<td>527</td>
<td></td>
</tr>
<tr>
<td>Randomized blocks error</td>
<td>8</td>
<td>323</td>
<td>40.4</td>
</tr>
</tbody>
</table>

This result has no practical use because no experiment should be planned if it has only 4 degrees of freedom for the estimate of the error variance. The estimated average standard error of the difference between two adjusted treatment effects is
\( \sqrt{6s^2/4} = 3.4 \). The adjusted treatment means are computed by adding \( \bar{Y} = 12.3 \) to each adjusted treatment effect, \( t'_{ij} = \left[ kT_{ij} - \bar{T}_{ij} \right] / 2k + \bar{Y} \) where \( \bar{T}_{ij} = (X_{1i} - Y_{1i} - X_{2j} + Y_{2j}) \). The \( \bar{T}_{ij} \) and \( t'_{ij} \) are

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-24</td>
<td>-6</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>-23</td>
<td>-5</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>-5</td>
<td>13</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>-52</td>
<td>2</td>
<td>50</td>
</tr>
</tbody>
</table>

The efficiency of this design as compared to randomized complete blocks is only

\[ 40.4 : 3(47.2)/2 = 0.57 \]

plus the fact we have only 4 instead of 8 error degrees of freedom.

If the simple lattice is duplicated several times so that \( 2d \) replications (2kd blocks) are used, the analysis is changed in the following respects:

(i) There are now \( (2d - 1) \) degrees of freedom for replications and \( 2d(k - 1) \) degrees of freedom for blocks in replications. The differences between block totals with the same treatments in the blocks are free of treatment effects and hence indicate real block differences. The block totals can be analyzed as follows for the X-group (and similarly for the Y group):

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>28.3</td>
<td>21.3</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>22.5</td>
<td>24.2</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
<td>21.5</td>
<td>38.2</td>
</tr>
<tr>
<td></td>
<td>75.3</td>
<td>65.3</td>
<td>81.4</td>
</tr>
</tbody>
</table>

Hence there are \( 2(d - 1)(k - 1) \) degrees of freedom for real block differences, and \( 2(k - 1) \) degrees of freedom with block and treatment effects mixed up.

(ii) Each \( X_{ij} \) and \( Y_{ij} \) is now a total of \( d \) yields and \( T_{ij} \) a total of \( 2d \) yields.

Hence all sums of squares must be divided by \( 2d \) instead of 2 (similarly for means). Also \( \bar{Y} \) will now be \( d \) instead of \( 1. \)

(iii) The above variances of the differences between two adjusted treatment effects are divided by \( d. \)

(iv) There are now \( (k - 1)(2dk - k - 1) \) degrees of freedom for error.
Most analyses of lattice designs now make use of a new computing technique first introduced by Yates for balanced designs (16). This technique utilizes the so-called inter-block information on treatment differences to improve the efficiency of the estimates. Since this analysis depends on the use of two sources of error, it will be discussed in Chapter 11 (Variance Components). One of the most widely circulated bulletins using the new method of analysis is reference (17). In this analysis, the sum of squares for blocks (adjusted for treatments) is computed. Then the sum of squares for treatments (adjusted for blocks) can be obtained by subtraction as indicated on page 204. The reader is advised to see reference (1) or (17) for the details of computing.

9.7.3. Other lattice designs. Yates (12) presents the theory for triple lattices. We shall not include the details in this book. The essential change is that each treatment is given a third subscript as was done for the Latin square design and a Z replication is introduced. This design is also presented in (17). If more than three replications are used, in almost all cases either the simple or triple lattice is duplicated or a balanced design is used.

It should be indicated that it is also possible to set up a lattice experiment in a Latin square design, called a lattice square design. The computing details are much more complicated than for the randomized blocks designs, but the reduction in error variance is often quite large especially for very heterogeneous experimental material. Details of these designs are also presented in (1).

Harshbarger (18) (19) introduced a new type of lattice design, called the rectangular lattice, in which the number of treatments is \( k(k + 1) \). This design enables the experimenter to utilize the simplicity of the lattice without restricting his experiments to exactly \( k^2 \) treatments. Cochran and Cox (1) present simplified computing techniques for this type of design also.
Exercise 9.7.1. The following data represent the yields (in bushels per acre, minus 30 bushels) of a 5 × 5 simple lattice experiment on soybean varieties with 4 replications. The variety numbers are given in parentheses.

<table>
<thead>
<tr>
<th>Replication I</th>
<th>Block Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) 6 (2) 7 (3) 5 (4) 8 (5) 6</td>
<td>32</td>
</tr>
<tr>
<td>(6) 16 (7) 12 (8) 12 (9) 13 (10) 8</td>
<td>61</td>
</tr>
<tr>
<td>(11) 17 (12) 7 (13) 7 (14) 9 (15) 14</td>
<td>54</td>
</tr>
<tr>
<td>(16) 18 (17) 16 (18) 13 (19) 13 (20) 14</td>
<td>74</td>
</tr>
<tr>
<td>(21) 14 (22) 15 (23) 11 (24) 14 (25) 14</td>
<td>68</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>289</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Replication II</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) 24 (6) 13 (11) 24 (16) 11 (21) 8</td>
<td>80</td>
</tr>
<tr>
<td>(2) 21 (7) 11 (12) 14 (17) 11 (22) 23</td>
<td>80</td>
</tr>
<tr>
<td>(3) 16 (8) 4 (13) 12 (18) 12 (23) 12</td>
<td>56</td>
</tr>
<tr>
<td>(4) 17 (9) 10 (14) 30 (19) 9 (24) 23</td>
<td>89</td>
</tr>
<tr>
<td>(5) 15 (10) 15 (15) 22 (20) 16 (25) 19</td>
<td>87</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>392</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Replication III</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) 13 (2) 26 (3) 9 (4) 13 (5) 11</td>
<td>72</td>
</tr>
<tr>
<td>(6) 15 (7) 18 (8) 22 (9) 11 (10) 15</td>
<td>81</td>
</tr>
<tr>
<td>(11) 19 (12) 10 (13) 10 (14) 10 (15) 16</td>
<td>65</td>
</tr>
<tr>
<td>(16) 21 (17) 16 (18) 17 (19) 4 (20) 17</td>
<td>75</td>
</tr>
<tr>
<td>(21) 15 (22) 12 (23) 13 (24) 20 (25) 8</td>
<td>68</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>361</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Replication IV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) 16 (6) 7 (11) 20 (16) 13 (21) 21</td>
<td>77</td>
</tr>
<tr>
<td>(2) 15 (7) 10 (12) 11 (17) 7 (22) 14</td>
<td>57</td>
</tr>
<tr>
<td>(3) 7 (8) 11 (13) 15 (18) 15 (23) 16</td>
<td>64</td>
</tr>
<tr>
<td>(4) 19 (9) 14 (14) 20 (19) 6 (24) 16</td>
<td>75</td>
</tr>
<tr>
<td>(5) 17 (10) 18 (15) 20 (20) 15 (25) 14</td>
<td>84</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>357</strong></td>
</tr>
</tbody>
</table>

(a) First analyze only the first two replications of this experiment, indicating if there are significant varietal effects, the average standard errors of adjusted varietal differences, and the efficiency of using the lattice design.

(b) Secondly analyze the entire experiment, making the necessary adjustments with d = 2.

(c) Show how this experiment might have been set out in the field.

Exercise 9.7.2. A randomized blocks experiment was set up to test the performance of 16 treatment combinations in two different blocks, each block having 4 rows with 4 treatments per row. The original set-up was a factorial experiment. It turned out that only two of the main treatments, designated as 1 and 0, were important. It also
became evident that the blocks were badly placed, since there were marked differences in fertility between the four rows, which ran across both blocks. In the analysis of the difference between the effects of L and 0, we are confronted with the difficulty that L and 0 did not each appear twice in each row of each block but sometimes L would appear 3 times and 0 once or vice versa. The yields are presented below:

<table>
<thead>
<tr>
<th>Row</th>
<th>Block I</th>
<th>Block II</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L</td>
<td>O</td>
<td>L</td>
</tr>
<tr>
<td>1</td>
<td>84,70, 81</td>
<td>66</td>
<td>63,97</td>
</tr>
<tr>
<td>2</td>
<td>146,171</td>
<td>148,137</td>
<td>189</td>
</tr>
<tr>
<td>3</td>
<td>247</td>
<td>179,218,228</td>
<td>195,189</td>
</tr>
<tr>
<td>4</td>
<td>177,153</td>
<td>123,166</td>
<td>145,141,130</td>
</tr>
<tr>
<td>Total</td>
<td>1129</td>
<td>1265</td>
<td>1149</td>
</tr>
</tbody>
</table>

|     | 2394 | 2250 |

(a) Set up the least squares equations to test if there was a real difference between L and 0 after adjusting for row effects.

(b) Show that the following analysis of variance is obtained:

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>648.00</td>
</tr>
<tr>
<td>Rows</td>
<td>70542.25</td>
</tr>
<tr>
<td>Treatments (adj.)</td>
<td>1313.21</td>
</tr>
<tr>
<td>Error</td>
<td>8006.04</td>
</tr>
</tbody>
</table>

(c) Fill in the appropriate degrees of freedom in (b) and make a test of the treatment differences.

**Exercise 9.7.3.** The factorial type of experiment can also be handled in an incomplete blocks design. Two different cultivation methods were used in the experiment presented in Example 9.5.1, giving a total of 12 treatments, but only 6 treatments were planted per block. The treatment effects can be divided as follows with the number of degrees of freedom per effect in parentheses: F(2), V(1), FV(2), C(1), FC(2), VC(1), FVC(2). The two cultivation methods (C) were allocated in the following order for each block (see the example for F and V):
\[
\begin{array}{ccccccc}
1a & 1b & 2a & 2b & 3a & 3b \\
2 & 1 & 1 & 2 & 1 & 2 \\
1 & 2 & 2 & 1 & 2 & 1 \\
1 & 2 & 2 & 1 & 1 & 2 \\
2 & 1 & 1 & 2 & 2 & 1 \\
1 & 2 & 1 & 2 & 2 & 1 \\
2 & 1 & 2 & 1 & 1 & 2 \\
\end{array}
\]

We note that there are three complete replications of the 12 treatments. The total yields of three plots for each of the 12 treatments were:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>111</th>
<th>112</th>
<th>121</th>
<th>122</th>
<th>211</th>
<th>212</th>
<th>221</th>
<th>222</th>
<th>311</th>
<th>312</th>
<th>321</th>
<th>322</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yield</td>
<td>411</td>
<td>444</td>
<td>561</td>
<td>657</td>
<td>479</td>
<td>578</td>
<td>659</td>
<td>645</td>
<td>451</td>
<td>517</td>
<td>546</td>
<td>601</td>
<td>6549</td>
</tr>
</tbody>
</table>

(a) Check that the above treatment totals are correct.

(b) Write out 11 independent effects, using \(F_X\) and \(F_q\) as before. Show that 9 of these are independent of block effects, and show that the sum of squares attributable to these 9 effects is 25,977.84.

(c) Show that \(VC, F_X VC\) and \(F_q VC\) are not independent of block effects; those effects are said to be confounded with block effects.

(d) Show that the sums of squares for the effects in (c) adjusted for blocks are 40.5, 154.1 and 129.6, respectively.

(e) Compute the new error mean square and make the necessary tests of significance.

(f) For further information on confounded factorial experiments, see (1), (8) and (9).

9.8.1. Assumptions made in the model. Perhaps it might be advisable to reemphasize
the assumptions which we have made in using the analysis of variance and the F-test
on experimental data. These assumptions and methods of handling aberrations from them
are treated in the March, 1947, issue of Biometrics (20). The computations as
summarized in the analysis of variance table are purely arithmetic devices to obtain
quantities which will add up to the total sum of squares. The following assumptions
must be met in order to apply our results to the solution of problems.

1. In order to connect the analysis of variance with the theory of linear regression,
we must assume that the various effects and errors are additive.

If the physical make-up of the data or of the operations producing the data are
of such nature that the effects are really not additive, then the sum of squares
attributable to such effects by the analysis of variance do not represent the true
effects. If non-additive effects are actually multiplicative, then we could use the
logarithms of the original data and have additive effects. Some other type of
transformation might be useful for other kinds of non-additivity. Except for the
multiplicative case, it is usually assumed that the additive assumption is a good
first approximation to the true relationship and that refinements here would produce
only minor improvements in the analysis.

2. If it is desired to use the analysis of variance for the purpose of estimating
the population variance or for making tests of significance, certain other assumptions
must be made.

   (a) The population residuals $\{\epsilon_i^2\}$ must be independent of one another and also
   independent of the main effects. In field experiments, it is known that the residuals
   in adjacent plots are usually positively correlated. A device, called randomization,
   is used to circumvent this difficulty. By this device, we allocate the treatments at
   random within a given block, so that there is an equal chance of any two treatments
   appearing in adjacent and in non-adjacent plots. Hence the expected value of the
total residuals for any one treatment is independent of that for any other treatment.
It should be emphasized that randomization does not remove the correlation between the inherent characteristics of adjacent plots; rather it provides a mechanism by which this expected correlation between two treatments tends to cancel with increased replication. A more complete treatment of this subject is given by R. A. Fisher (9), Cochran (20), and Yates (21). Cochran (1) makes the following remarks concerning randomization: "Randomization is somewhat analogous to insurance, in that it is a precaution against disturbances that may or may not occur and that may or may not be serious if they do occur. It is generally advisable to take the trouble to randomize even when it is not expected that there will be any serious bias from failure to randomize." (p. 8)

Such data as daily prices and production figures cannot fulfill the randomization requirement. Hence it is frequently stated that the analysis of variance is not a valid method of testing for yearly or monthly price differences. To date, the mathematical difficulties of testing for the existence of correlations between residuals have prevented any definite statements on the validity of the analysis of variance for such data. Research in the field of serial correlation may lead to an approximate solution of this problem. The serial correlation coefficient measures the correlation between successive members of a series. See reference (22) for a recent article on this subject and for a rather complete bibliography to date.

(b) The residuals are supposed to be normally distributed with the same variance from plot to plot. It has been stated that the effect of slight deviations from normality is not too serious but this may not be true for significance levels with low probabilities ($P = .01$, for example). It has been shown that the $t$-test is upset when the variances from treatment to treatment are not equal. The same conclusion holds for the $F$-test. Under some conditions, a transformation of the data will also improve the agreement between these assumptions and reality (20).
9.8.2. **Principles of setting up an experiment.** We have made numerous references to the recent book by Cochran and Cox on Experimental Designs (1) because this book incorporates most of the present concepts of the principles of setting up and analyzing experiments. Since many readers of our text may not have an opportunity to read this extensive book on designs, we will include here some general principles of setting up experiments. For a discussion of the simpler designs and their analysis, the reader should be sure to read the chapters on the analysis of variance by G. W. Snedecor (23).

(i) **The experimenter should clearly set forth his objectives before proceeding with the experiment.** Is this a preliminary experiment to determine the future course of experimentation or is it intended to furnish answers to immediate questions? Are the results to be carried into practical use at once or are they to be used to explain aspects of theory not adequately understood before? Are you mainly interested in estimates or in tests of significance?

(ii) **The experiment should be described in detail.** The treatments should be clearly defined. Is it necessary to use a control treatment in order to make comparisons with past results? The size of the experiment should be determined. If insufficient funds are available to conduct an experiment from which useful results can be obtained, the experiment should not be started. And above all, the necessary material to conduct the experiment should be available.

(iii) **An outline of the analysis should be drawn up before the experiment is started.**

9.8.3. **Methods of increasing the accuracy of the experiment.** Anyone who has attended a lecture by Professor Gertrude Cox on Experimental Design will recognize the remarks that follow. **Accuracy** refers to the success of estimating the true value of a quantity; it is often confused with **precision**, which refers to the clustering of sample values about their own average, which will not be the true value if this
average is biased. Precision can be thought of as the inverse of variance. Hence accuracy is a more inclusive term, since it involves both unbiasedness and precision. Often the experimenter has to choose between an unbiased estimate with rather low precision and a slightly biased one with high precision. The choice of the proper estimate is often dictated by circumstances beyond his control, but certain methods of increasing the accuracy of the experiment should be kept in mind. If there is no bias in the experimental procedure, accuracy and precision are for all practical purposes synonymous.

(i) **Accuracy can generally be increased by increasing the size of the experiment.** There are certain limitations to this statement, such as the fact that increasing size may bring in more heterogeneous material and also may result in a poorer supervision of the experiment with a possible biased result. The latter point is often true in industrial and psychological experiments and in sample surveys. But in general the accuracy of an estimate increases with increasing size of the experiment. Increasing the number of replications or treatments also furnishes more degrees of freedom for the estimate of the experimental error.

(ii) **The experimental techniques should be refined as much as possible.**

(a) There should be a uniform method of applying the treatments to the experimental units.

(b) Sufficient control over external influences should be exercised so that every treatment operates under as near the same conditions as possible.

(c) Unbiased measures of the treatment effects should be devised so that they are fully understood by those running the experiment and by other research workers. As yet we do not have good enough measures of such things as socio-economic status, educational progress and economic conditions to enable us to compare adequately the results of one experiment or sample with another.
(d) As far as possible, checks should be set up to avoid gross errors in recording and analyzing the data.

(iii) The experimental material should be selected to suit the experiment.

(a) The size and shape of the experimental unit should be prepared to achieve maximum accuracy and unbiasedness.

(b) Often additional measurements can be taken to help explain the final results (see Chapter 10 on covariance techniques).

(c) Finally the treatments should be grouped together in the best manner. In other words, the proper selection of the experimental design is of utmost importance. If there are too many treatments or the experimental units are quite heterogenous, an incomplete blocks design might be used. If interactions are important, some type of a factorial design is needed; if higher order interactions are not important, some system of confounding might be used. The problem of whether to balance the treatments or not must be decided on the basis of the importance of various comparisons and the amount of experimental material available.

2.3.4. Construction of experimental designs. In all of the theoretical discussions in the previous sections of this paper, we have assumed that the design was known and proceeded to set up an analysis for this design. The reader might be interested in knowing how to construct designs in the first place so that the analysis will be relatively simple.

(a) There is no difficulty in setting up completely randomized or randomized complete blocks or complete Latin square designs, except to remember that randomization is necessary.

(b) In planning confounded factorial designs, the main principle is to restrict the confounding to high order interactions, if this is possible. One principle must always be remembered: if two interactions are confounded, then a third effect is also confounded - this third effect is formed by casting out all like letters
in the first two. For example if ABC and ABD are confounded, CD will be also.
Hence it is often necessary to adjust the confounding so as to protect main
effects and two-factor interactions. We will present a few examples of how to
construct confounded factorials. See Yates (8) for the details on other
examples.

Example 2.8.1. If a design 4 is of the 2^n character (n factors each at two levels),
the construction of a confounded design is simplified by the use of the (+,−) system.
For example if we wish to use 2^4 (= 16) treatments in 4 blocks of 4 treatments each,
we are led to confound 3 degrees of freedom (the number of degrees of freedom between
blocks). Let us designate the 4 factors as A, B, C, and D with 0 standing for the
low level and 1 the high level of a factor. We cannot confound on ABCD, because if
we select any three factor interaction, we will automatically confound a main effect
by the above rule. Hence we consider confounding two of the four three-factor
interactions and one two-factor interaction, e.g., ABC, ABD, and CD. We set up the
(+,−) system for each of these three effects as follows.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>ABC</th>
<th>ABD</th>
<th>CD</th>
<th>Block</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>x</td>
</tr>
<tr>
<td>1000</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>x</td>
</tr>
<tr>
<td>0100</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>x</td>
</tr>
<tr>
<td>1100</td>
<td>−</td>
<td>−</td>
<td>±</td>
<td>x</td>
</tr>
<tr>
<td>0010</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>x</td>
</tr>
<tr>
<td>1010</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>x</td>
</tr>
<tr>
<td>0110</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>x</td>
</tr>
<tr>
<td>1110</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>x</td>
</tr>
<tr>
<td>0001</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>x</td>
</tr>
<tr>
<td>1001</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>x</td>
</tr>
<tr>
<td>0101</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>x</td>
</tr>
<tr>
<td>1101</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>x</td>
</tr>
<tr>
<td>0011</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>x</td>
</tr>
<tr>
<td>1011</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>x</td>
</tr>
<tr>
<td>0111</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>x</td>
</tr>
<tr>
<td>1111</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>x</td>
</tr>
</tbody>
</table>

We note that there is a * for the treatment having the high level of all the
factors in a particular interaction, − for one low level and rest high levels, etc.
Also CD = (ABC)(ABD). The treatments with the same combination of (+,−) signs are
assigned to the same block. This method can be used for all 2^p designs.
Example 9.8.2. Consider a $3^2$ design with the levels designated as 0, 1, and 2. The 9 treatments can be arranged in a $3 \times 3$ table.

\[
\begin{array}{ccc}
00 & 01 & 02 \\
10 & 11 & 12 \\
20 & 21 & 22 \\
\end{array}
\]

The $AB$ interaction (4 degrees of freedom) can be split into two components, each with 2 degrees of freedom. We compute

\[
I_0 = (00 + 11 + 22), \quad I_1 = (01 + 12 + 20), \quad I_2 = (02 + 10 + 21)
\]

\[
J_0 = (00 + 12 + 21), \quad J_1 = (01 + 10 + 22), \quad J_2 = (02 + 11 + 20)
\]

The two components are

\[
I = \frac{\sum I_i^2}{3} - \frac{G^2}{9}, \quad J = \frac{\sum J_i^2}{3} - \frac{G^2}{9}
\]

If we wanted to use only 2 treatments per block, we could confound either the $I$ or $J$ part of the $AB$ interaction by planting the $I$ or $J$ combination of treatments in blocks. We note that these combinations do not confound the main effects, since all 3 levels of each factor are present in each combination. See (8) for the rules when $n \geq 2$.

(c) References (13) and (24) present methods of constructing balanced incomplete blocks designs. The construction of the lattice designs is quite simple, since the experimenter needs only to randomize the treatment numbers, then the blocks within a replication and the treatments in each block. We will present one example of a non-lattice.

Example 9.8.3. Consider the design for the 7 treatments in blocks of 3 in Exercise 9.6.3. There are $7C_3 = 35$ combinations of these 7 treatments in groups of 3, but if we restrict ourselves to enough combinations so that each treatment appears once and only once with every other treatment ($\lambda = 1$), we can get by with 7 blocks. The method of construction is one of gradual elimination. Obviously there are many ways of selecting the set of 7 from the set of 35 combinations. One such set of 7 is

\[
ABC, \ ADE, \ AFG, \ BDF, \ BEG, \ CDG, \ CEF
\]
Another was used in Exercise 9.6.3. Probably the best procedure to use is to select some basic rule, such as the one given in (1).

\[ \text{ABD, BCE, CDF, DEG, EFA, FGB, GAC} \]

by cyclic substitution (the first letter is one removed from the second and the third two removed from the second, where \( F + 2 = A \), for example). Then randomly assign your 7 treatments to the 7 letters.

(d) Kempthorne and Federer (25) present the general theory of prime-power lattices.


I. \( p^n \) Varieties in Blocks of \( p \) Plots. *Biometrics* 4:54-79 (1948).

II. \( p^n \) Varieties in Blocks of \( p^k \) Plots and in Squares. *Biometrics* 4:109-121 (1948).


**OTHER REFERENCES**


7. Irwin, J. O. "Mathematical Theorems Involved in the Analysis of Variance." 

8. Irwin, J. O. "Independence of the Constituent Items in the Analysis of Variance." 


11. Li, J. Design and Statistical Analysis of Some Confounded Factorial Experiments, 


    The Commercial Press, Ltd., Shanghai (1936).

    Univ. Puerto Rico (1943).

    New York (1937).

    New York (1937).


18. "Student" (W. S. Gosset) "The Probable Error of a Mean." Biometrika 6:1-25 
    (1908).

19. Weiss, M. C. and Cox, G. M. Balanced Incomplete Block and Lattice Square Designs 
    for Testing Yield Differences among Large Numbers of Soybean Varieties. 

20. Welch, B. L. "On the z-test in Randomized Blocks and Latin Squares." 


    Soc. 1:26-61 (1934).


CHAPTER 10
THE ANALYSIS OF COVARIANCE

10.1. Introduction.

In chapter 9, we considered various types of experimental designs to estimate treatment effects and to test for differences among these treatments. Frequently the experimenter wishes to make these estimates and tests on some dependent variable after adjusting for the effects of one or more independent variables. For example, he might wish to test the effectiveness of various rations on the gains in weight of hogs after adjustments have been made for the initial weights of the hogs. Other independent variates may be used in order to study the effect of the rations adjusted for the effects of such external factors as temperature of the pens or sunlight. An agronomic experiment might be improved by adjusting crop yield for weather and soil conditions or for unequal stand. In an educational experiment to test for differences between teaching methods, it might be advisable to adjust the results for the mental age or for some test score secured before the experiment started. Cochran has discussed theoretical and practical aspects of the analysis of covariance in (1) and (2). R. A. Fisher first introduced the method in (3).

10.2. The Use of Simple Covariance for a Randomized Blocks Experiment.

10.2.1. We shall not attempt to present the theory of covariance for all of the experimental designs of Chapter 9, since it is believed that the method can be adequately demonstrated with the randomized complete blocks type of experiment. As in section 9.3, we shall assume that the experimenter has \( p \) treatments, each assigned at random to an individual plot in each of \( b \) blocks. The variable to be estimated is designated as \( Y \), and the independent variable for which \( Y \) is to be adjusted is designated as \( X \). The experimental model for the yield of the \( i \)th treatment in the \( j \)th block is

\[
Y_{ij} = \mu + \tau_i^* + \beta_j^* + \beta x_{ij} + \epsilon_{ij} = \mu + t_i^* + b_j^* + bx_{ij} + e_{ij},
\]

where \( x_{ij} = X_{ij} - \bar{X} \), and \( \tau_i^* \) and \( \beta_j^* \) are the treatment and block effects adjusted.
for the effect of the X's. Since \( Sx = 0 \), it is obvious that \( \mu^* = \mu \); hence, we shall use \( \mu \) and its estimate, \( \bar{m} \), in the theory which follows. As usual \( \sum i_i \beta^*_i = 0 \).

The X's are assumed to be fixed, and hence not influenced by the treatments. In case X has some effect on the yield (Y) and the experimenter was not able to maintain the same value of X for all treatments, it is desirable to estimate the treatment effects after the yields have been adjusted for the effect of X. If the values of X are actually influenced by the treatments but can be measured without error, an analysis of covariance can still be run, but interpretations are often quite difficult. As indicated in Chapter 8, if we estimate \( \mu, \tau^*_i, \beta^*_j, \) and \( \beta \) by the method of least squares, each of the first three estimates will be adjusted for the linear effect of X. It should be emphasized here that we are here considering only the linear effect of X; however, any other function of X could be used as the independent variate, and we might consider several independent variates in a multiple covariance.

10.2.2: The least squares equations for \( m, t^*_i, b^*_j \) and \( b \), respectively, are:

\[
\begin{align*}
\text{rpm} &= SY \\
\text{rm} + rt^*_i + bx^*_i &= T_i \\
pm + pb^*_j + bx^*_j &= B_j \\
\sum_i t^*_ix^*_i + \sum_j b^*_jx^*_j + bSx^2 &= SxY = Sxy,
\end{align*}
\]

where \( x^*_i = \frac{Sx_i}{Sx} = x_i - \bar{x}, \) \( x^*_j = \frac{Sx_j}{Sx} = x_j - \bar{x}, \) and \( T_i \) and \( B_j \) were defined in Section 9.3. In other words, \( x^*_i \) and \( x^*_j \) are treatment and block sums for X. It is obvious that \( \sum_i x^*_i = \sum x^*_j = 0 \).

The solutions to the least squares equations are:

\[ m = SY/rp = \bar{Y} \]
\[ b = \frac{S_{xy} - p \sum_j \bar{y}_j \bar{x}_j - r \sum_i \bar{t}_i \bar{x}_{i.}}{S_x^2 - p \sum_j \bar{x}_j^2 - r \sum_i \bar{x}_{i.}^2} \frac{E_{xy}}{E_{xx}} \]

\[ t^*_i = (\bar{T}_1 - \bar{Y}) - b \bar{x}_{i.} = t_i - b \bar{x}_{i.} \]

\[ b^*_j = (\bar{B}_j - \bar{Y}) - b \bar{x}_{j.} = b_j - b \bar{x}_{j.} \]

where \( E_{xy} \) and \( E_{xx} \) are similar to SSE but applied to \( x^2 \) and \( xy \), \( \bar{x}_{i.} = x_{i.} / r \), \( \bar{x}_{j.} = x_{j.} / p \), and similarly for \( T \) and \( B \).

We see that an adjusted treatment effect \((t^*)\) is estimated by subtracting an adjustment factor from the unadjusted effect \((t)\). The adjustment factor, \( b \bar{x}_{i.} \), is simply the average change in \( Y \) for a unit change in \( X \) multiplied by the difference between the treatment mean of \( X \) and \( \bar{X} \) \((\bar{x}_{i.} = \bar{x}_{i.} - \bar{X})\). Hence each treatment effect is adjusted to the average effect expected if all treatments had operated with the mean value of \( X \).

It can be shown that

\[ E_{xy} = S((x_{i.} - \bar{x}_{i.} - \bar{x}_i)^2(\bar{y}_{1.} - \bar{B}_j - \bar{T}_1)) \]

\[ E_{xx} = S((x_{i.} - \bar{x}_{i.} - \bar{x}_i)^2. \]

10.2.2. The estimates in 10.2.2 can be put in terms of the parameters and \( \{ \xi_{ij} \} \) as follows:

\[ \bar{y} = \mu + \bar{e} \]

\[ y_{ij} = \bar{t}_{i.} + \beta^* j + \beta x_{ij} + \xi_{ij} - \bar{e} \]

\[ \bar{B}_j = \mu + \bar{t}_{j.} + \beta \bar{x}_{j.} + \bar{e}_{j.} \]

\[ \bar{T}_i = \mu + \bar{t}_{i.} + \beta \bar{x}_{i.} + \bar{e}_{i.} \]
\[ b = \frac{E_{xy}}{E_{xx}} = \beta + \frac{S(x_{ij} - \bar{x}_{.j} - \bar{x}_{i.})(\bar{e}_{ij} - \bar{\bar{e}}_{.j} - \bar{\bar{e}}_{i.} - \bar{\bar{e}})}{E_{xx}} \]

\[ = \beta + \frac{S(x_{ij} - \bar{x}_{.j} - \bar{x}_{i.}) \bar{e}_{ij}}{E_{xx}} + \bar{\bar{e}}_b \]

\[ t_i^* = T_i^* + \bar{\bar{e}}_{i.} - \bar{\bar{e}} - \bar{x}_{i.} \in b \]

\[ b_j^* = \beta_j^* + \bar{\bar{e}}_{.j} - \bar{\bar{e}} - \bar{x}_{.j} \in b \]

where \( \bar{\bar{e}} = s \bar{e}/r_p \), \( \bar{\bar{e}}_{.j} = s \bar{e}_{.j}/p \), and \( \bar{\bar{e}}_{i.} = s \bar{e}_{i.}/r \). Hence \( \bar{\bar{t}}_i^* \), \( b \), \( t_i^* \), and \( b_j^* \) are unbiased estimates of \( \mu \), \( \beta \), \( T_i^* \), and \( \beta_j^* \), respectively.

An unbiased estimate of an adjusted treatment mean is

\[ t_i^* + \bar{\bar{t}} = \bar{\bar{t}}_i^* + \mu + \bar{\bar{e}}_{i.} - \bar{x}_{i.} \in b \]

Similarly the estimate of the difference between two treatment means is

\[ d^* = t_i^* - t_j^* = (T_i^* - T_j^*) + (\bar{\bar{e}}_{i.} - \bar{\bar{e}}_{j.}) = (\bar{x}_{i.} - \bar{x}_{j.}) \in b \]

The variance of the difference between two treatment means is

\[ \sigma^2(d^*) = \sigma^2 \left[ \frac{2}{r} + \frac{(\bar{x}_{i.} - \bar{x}_{j.})^2}{E_{xx}} \right] \]

In computing this variance, we use \( \bar{x}_{i.} - \bar{x}_j = \bar{x}_{i.} - \bar{x}_j \). It might be noted that in computing this variance, we computed \( \sigma^2(b) = \sigma^2/E_{xx} \).

**10.2.4.** From least squares theory, we know that the residual sum of squares is given by

\[ SSE^* = Sy^2 - \sum_i t_i^* T_i - \sum_j b_j^* B_j - bS_{xy} \]

\[ = Sy^2 - \frac{1}{r} \left[ \sum_i T_i^2 - rc \right] - \frac{1}{p} \left[ \sum_j B_j^2 - pc \right] - bE_{xy} \]
where \( C = \frac{(SY)^2}{rp} \) and \( Sy^2 = SY^2 - C \). But this is the usual error sum of squares for the analysis of variance (9.3.4)

\[
SSE = Sy^2 - SSB - SST
\]

minus the reduction due to regression when \( x \) and \( y \) are adjusted for block and treatment effects.

The added reduction due to treatments is found by omitting the treatment constants from the model equation. The new residual variance is

\[
(SSE^*)' = Sy^2 - SSB - b'Exy,
\]

where

\[
b^* = \frac{Exy}{Exx} = \frac{S(x_{i,j} - \bar{x}_j)(y_{i,j} - \bar{y}_j)}{S(x_{i,j} - \bar{x}_{*j})^2}
\]

This is simply \( SSE' = Sy^2 - SSB \) minus the reduction due to regression when \( x \) and \( y \) are adjusted for block effects only.

The added reduction due to treatments is

\[
SST^* = (SSE^*)' - SSE^*
\]

Under the null hypothesis \( \sum \alpha_i = 0 \), \( SST^* \) was shown in Chapter 8 to be distributed as \( \chi^2 \sigma^2 \) with \( (p - 1) \) degrees of freedom and independently of \( SSE^* \), which is distributed as \( \chi^2 \sigma^2 \) with \( (rp - r - p) \) degrees of freedom. Hence

\[
F = \frac{SST^*/(p - 1)}{SSE^*/(rp - r - p)} = \frac{MSST^*}{s^*2}
\]

with \( (p - 1, rp - r - p) \) degrees of freedom can be used to test the above null hypothesis.\(^1\) Hence

\[
\sigma^2 = SSE^*/(rp - r - p) = s^*2
\]

This estimated variance can be used to set up confidence limits for differences between adjusted treatment means.

\(^1\) \( s^*2 \) is usually designated as \( s^2_{y \cdot x} \).
10.2.5. The analysis of covariance table is

<table>
<thead>
<tr>
<th>Source</th>
<th>Degrees of Freedom</th>
<th>Original Sums</th>
<th>Adjusted Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$y^2$</td>
<td>$xy$</td>
</tr>
<tr>
<td>Total</td>
<td>$rp - 1$</td>
<td>$Sy^2$</td>
<td>$Sxy$</td>
</tr>
<tr>
<td>Blocks</td>
<td>$r - 1$</td>
<td>$SSE$</td>
<td>$B_{xy}$</td>
</tr>
<tr>
<td>Treatments</td>
<td>$p - 1$</td>
<td>$SST$</td>
<td>$T_{xy}$</td>
</tr>
<tr>
<td>Error</td>
<td>$(r - 1)(p - 1)$</td>
<td>$SSE'$</td>
<td>$E_{xy}'$</td>
</tr>
<tr>
<td>(Error)$'$</td>
<td>$r(p - 1)$</td>
<td>$SSE'$</td>
<td>$E_{xy}'$</td>
</tr>
</tbody>
</table>

Adjusted Treatments = (Error)$'$ - (Error) $= p - 1$ $SST^*$ $MST^*$

In this analysis,

$E_{xy} = p \sum_j B_{j} X_j \bar{x} = 1 \sum_j B_{j} X_j \bar{x} - \frac{(SY)(SX)}{rp}$,

$T_{xy} = r \sum_i I_{x_i} X_i = \frac{r}{I} - \frac{(SY)(SX)}{rp}$,

and similarly $B_{xx}$ and $T_{xx}$ are block and treatment sums of squares for $x$.

10.2.6. The analysis of covariance has two main uses:

(i) to reduce the error variance by eliminating the plot-to-plot variation attributable to fluctuations in the independent variate,

(ii) to eliminate any bias in treatment comparisons caused by an uneven distribution of the independent variate to the various treatments.

The efficiency of covariance in reducing the error variance is given by the ratio of the variance of the difference between two unadjusted treatment means ($2s^2/r$) to the average variance of the difference between two adjusted treatment means. Finney (4) shows that the average variance of the difference between two adjusted means is

$$s^2(d^*) = \frac{2s^2}{r} \left[ 1 + \frac{T_{xx}}{(p - 1)E_{xx}} \right].$$

Hence $I = s^2/s^2 \left[ 1 + \frac{T_{xx}}{(p - 1)E_{xx}} \right]$. 
It is rather difficult to assess the effectiveness of covariance in eliminating the effect of X on the treatment means. It has sometimes been stated that one might test for treatment differences in X by use of

\[ F_X = (r - 1) \frac{T_{XX}}{E_{XX}} \]

If \( F_X \) is not significant, the experimenter is told that he can attribute adjusted yield differences to the treatments; however, if \( F_X \) is significant, he is advised to be cautious because adjusted yield differences might be attributed to differences in X. Actually the experimenter should consider whether treatment differences in X were inherent in the treatments (such as poor germination resulting in a low stand) or were the result of external circumstances. If the latter is the case, then covariance should be used to eliminate a bias in estimating treatment differences. But if the treatments actually produce differences in X, the experimenter should take this into account in making recommendations.

**Example 10.1.** Snedecor (5) presents an example of the analysis of covariance of yield of sugar beets (Y) adjusted for stand (X). The yields in tons per acre and the stand in numbers of beets per plot are presented on page 247. The computations are as follows:

\[
\begin{align*}
S_{XY}^2 &= 3,587,590 \\
S_{XY} &= 67,664.27 \\
S_Y^2 &= 1316.1479 \\
\frac{(sx)^2}{42} &= 3,435,432 \\
\frac{(sx)(sy)}{42} &= 63,500.58 \\
\frac{(sy)^2}{42} &= 1173.7457 \\
S_X^2 &= 152,158 \\
S_X &= 4,163.69 \\
S_Y &= 142,4022 \\
E_X &= 4,163.69 - \left[ \frac{(2015)(35.00) + \cdots + (2106)(29.35)}{7} - \frac{(1378)(20.12) + \cdots + (2121)(44.12)}{6} \right] \\
&= 4,163.69 + 116.56 - 3,598.05 = 662.20. \\
E_{XX} &= 28,665.10; b = \frac{E_{XY}}{E_{XX}} = 0.023799. \\
SSE* &= S_Y^2 - SSB - SST - bE_{XY} \\
&= 142,4022 - 6,3134 - 112,8562 - 16,2357 = 6,9969. 
\end{align*}
\]
The results are presented in the following analysis of covariance table.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>D.F.</th>
<th>$S_y^2$</th>
<th>$S_{xy}$</th>
<th>$S_x^2$</th>
<th>Adjusted Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S.S.</td>
</tr>
<tr>
<td>Total</td>
<td>41</td>
<td>1424.022</td>
<td>4.163.69</td>
<td>152,158.00</td>
<td></td>
</tr>
<tr>
<td>Blocks</td>
<td>5</td>
<td>6.33.34</td>
<td>-1.316.56</td>
<td>7,472.57</td>
<td></td>
</tr>
<tr>
<td>Treatments</td>
<td>6</td>
<td>112,8562</td>
<td>3,708.05</td>
<td>116,020.33</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>30</td>
<td>23,2226</td>
<td>682.20</td>
<td>28,665.10</td>
<td>6.9969</td>
</tr>
<tr>
<td>(Error)'</td>
<td>36</td>
<td>156,0888</td>
<td>4,280.25</td>
<td>144,685.43</td>
<td>9.4555</td>
</tr>
<tr>
<td>Adjusted Treatments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.4686</td>
</tr>
</tbody>
</table>

$F = 0.4114/0.2413 = 1.70 \quad \text{P} > .10$

Since $F$ is not significant, we conclude that the treatments did not differ in their mean yields after adjusting for stand. This experiment falls in the uncertain class, because the stand is a result of the experiment and hence may be a treatment characteristic. For example a given treatment may produce good germination or a good start on the part of the plant so that its main contribution to yield may be in producing a good stand. If this is so, an adjustment for stand will cancel out the treatment effects. A test of $T_{XX} = 116, 020.33$ against $E_{XX} = 28,665.10$ gives
\( F_x = 20.24 \), a highly significant value. This indicates that there were actually differences in stand from one treatment to another. Furthermore we note that a test of the treatment effects on yield not adjusted for stand gives

\[
F = \frac{112.8562/23.2326}{6/30} = 24.29,
\]
also a highly significant value. Hence we conclude that there were definite treatment effects on yield but that these effects probably were arrived at indirectly through different stands, which then resulted in different yields. One further complication might be mentioned. Yields for low stand are often higher per plant than for high stand because of less competition for the available plant food and moisture. This apparently did not happen in our sugar beet example, but in general should be considered. Also it should be noted again that we only considered the linear effect of stand on yield; often the effect may be curvilinear. A final point is that although the treatments did affect the stand, the stand could be measured without error.

In this case, there is little need to set up the average variance of adjusted treatment differences, but we present it to complete the example:

\[
es^2(a^*) = \frac{0.4826}{6} (1 + \frac{19.337}{28.665}) = 0.135.
\]

The exact variance for \( (t_1^* - t_2^*) \), for example, is

\[
0.2413 \left[ \frac{2}{6} + \frac{(66)^2}{28.665} \right] = 0.117.
\]

The unadjusted and adjusted treatment means are

<table>
<thead>
<tr>
<th>Treatment</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjusted</td>
<td>4.693</td>
<td>5.399</td>
<td>4.931</td>
<td>5.828</td>
<td>5.526</td>
<td>4.878</td>
<td>5.746</td>
<td>5.286</td>
</tr>
</tbody>
</table>

The variance of an unadjusted mean would have been

\[
\left( \frac{2}{6} \right) \left( \frac{23.2326}{30} \right) = 0.258.
\]

Hence \( I = \frac{0.258}{0.135} = 191\% \).
10.3. The Use of Simple Covariance for Other Experimental Designs.

No attempt will be made in this book to outline in detail the computational procedures for the use of simple covariance with other experimental designs.

The computing procedure is exactly the same as for randomized blocks, to wit

(i) Set up the usual analysis of variance given in Chapter 9, but include the sum of squares for $x$ and the sum of cross-products as well as the sum of squares for $y$ for each line in the analysis.

(ii) Compute the "Error" line by subtraction for $\text{SSE}, \text{E}_{xy}$ and $\text{E}_{xx}$.

(iii) Compute $\text{SSE}^* = \text{SSE} - b \text{E}_{xy} = \text{SSE} - \frac{E_{xy}^2}{E_{xx}}$.

(iv) Compute the $(\text{Error} + \text{Treatment}) = (\text{Error})'$ line by adding the "Treatment" line to the "Error" line.

(v) $(\text{SSE}^*)' = \text{SSE}' - b' \text{E}_{xy}^*$.

(vi) Compute the adjusted sum of squares for treatments, $\text{SST}^* = (\text{SSE}^*)' - \text{SSE}^*$.

$\text{MST}^* = \frac{\text{SST}^*}{(p - 1)}$

(vi') $F = \frac{\text{MST}^*}{s^*2}$

(vii) $t_{i}^* + \bar{y} = \bar{y}_{i} - b(\bar{x}_{i} - \bar{x})$

(ix) $s^2(\bar{y}) = \frac{2s^*2}{r} [1 + \frac{T_{xx}}{(p - 1)E_{xx}}]$

If the experimenter wishes to test a single degree of freedom of the $(p - 1)$ treatment degrees of freedom, he must go through the same procedure as outlined above but with this one degree of freedom replacing the "Treatment" line in the computations. If several of these single components are desired, an approximate method of computation has been suggested by Cochran and Cox (1). Compute

\[ S(y - bx)^2 = S y^2 - 2bs_{xy} + b^2s_x^2 \]

for each component of the treatment sum of squares. This computation would give the correct value of $\text{SSE}^*$ but slightly over-estimates each of the components of $\text{SST}^*$.

However the bias is generally small, and if the experiment is a complicated factorial, much labor is saved in testing the various main effects and interactions.

See reference (6) for the use of covariance with lattice designs.
10.4. The Use of Multiple Covariance: \( Y_{ij} = \mu + \tau_i^* + \beta_j^* + \sum_{k=1}^{m} \beta_k x_{kij} + \epsilon_{ij} \)

Snedecor (5) presents an example of a randomized block experiment on wheat yields in Great Britain with two independent variates (height of shoots at ear emergence and number of plants at tillering). The treatments were 6 different places and the blocks were 3 different years. The computing procedure is as follows for \( m \) independent variates:

(i) Derive a table of sums of squares and cross-products for Total, Blocks and Treatments (and any other component in the analysis, such as columns in a Latin Square).

(ii) Compute the "Error" line for each sum of squares and cross-products by subtracting Blocks, Treatments and any other component from the Total. Use these "Error" values to compute the \( \{ b_k \} \) and \( R_0^2 \), the squared multiple correlation coefficient for the "Error" line.

(iii) Compute an (Error + Treatment) line and \( R_{0+1}^2 \).

(iv) \( SSE^* = SSE - R_0^2 SSE = SSE(1 - R_0^2) \)

\((SSE^*)^* = SSE^*(1 - R_0^2 + t)\)

(v) \( s^* = \sqrt{SSE^*} \) (Error D.F. = number of independent variates)

\( MST^* = \left[ (SSE^*)^* - SSE^* \right] / (p - l) \)

\( F = MST^*/s^* \)

(vi) \( t_i^* + \bar{Y} = T_1 - \sum_{k=1}^{m} b_k(\bar{x}_{ki.} - \bar{x}_k) \),

where \( b_k \) is estimated in (ii).

(vii) If estimates of the variances of adjusted treatment differences are wanted, the b's should be derived by a matrix-inversion method in order to obtain the variances and covariances of the b's.

\[ s^2(t_i^* - t_j^*) = s^* \left[ \frac{2}{r} + \sum_{k} \sum_{k'} c_{kk'} (\bar{x}_{ki.} - \bar{x}_k) (\bar{x}_{kj.} - \bar{x}_k) \right], \]

where \( c_{kk'} \) is the element in the inverse matrix of the "Error" line.
The average variance of adjusted treatment differences is

\[
\frac{2s^2}{r} \left[ 1 + \sum_k \sum_{k'} \frac{\bar{c}_{kk'} T_{kk'}}{(p-1)} \right],
\]

where \( T_{kk'} \) is the sum of cross-products for \((x_{ki}, x_{k'i})\) in the "Treatment" line of (1).

\[
T_{kk'} = \frac{s(x_{ki} - \bar{x}_i)(x_{k'i} - \bar{x}_{k'i})}{r}.
\]

**Exercise 10.1.** Derive these relationships for \(E_{xy}\) and \(E_{xx}^2\):

\[
E_{xy} = s(x_{ij} - \bar{x}_i - \bar{x}_j)(y_{ij} - \bar{r}_j - \bar{T}_i)
\]

\[
E_{xx} = s(x_{ij} - \bar{x}_i - \bar{x}_j)^2
\]

**Exercise 10.2.** (a) Derive Finney's result for the average variance of the difference between two adjusted treatment means, presented in section 10.2.6.

(b) Also, our result for multiple covariance given in section 10.4.

**Exercise 10.3.** (a) Make a complete least squares solution of a simple covariance analysis for a Latin Square design.

(b) Illustrate the results in (a) by analyzing the following 5 x 5 Latin Square experiment on the yield in bags per acre of No. 1 Irish potatoes \((Y)\), adjusted for the percentage of No. 1's \((X)\). The experiment was conducted by W. L. Nelson of the North Carolina Agricultural Experiment Station on the H. M. Lewis farm in 1945. The treatments were different amounts (lbs.) of \(P_2O_5\) per acre: \(a = 0\), \(b = 40\), \(c = 80\), \(d = 120\), \(e = 160\).

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows</td>
<td>t</td>
<td>Y</td>
<td>X</td>
<td>Y</td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>Y</td>
<td>X</td>
<td>Y</td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>Y</td>
<td>X</td>
<td>Y</td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>Y</td>
<td>X</td>
<td>Y</td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>Y</td>
<td>X</td>
<td>Y</td>
<td>X</td>
<td>Y</td>
</tr>
</tbody>
</table>

\[
S_{Y} = 595.038.38 \quad S_{XY} = 351.944.8 \quad S_{X} = 210.085
\]

Show that \(b = 2.0616\) and \(a^2 = 127.18\).
(c) Analyze the linear component of the treatment mean square by the exact method and by the Cochran and Cox approximation presented in Section 10.3. Then show that the deviations from a linear trend are not significant.

**Exercise 10.4.** Consider a single degree of freedom to test for the adjusted effect of phosphate in the sugar beet example (10.1): \( t = t_1 + t_2 - t_3 + t_4 - t_6 + t_7 \).
(a) Use the exact procedure to show that the effect of phosphate is significant even after adjusting for stand.
(b) Show that the approximate procedure presented in Section 10.3 cannot be used in this case.

**Exercise 10.5.** The vitamin B₂ example used in Chapter 8 and presented on page 116 can also be analyzed by covariance. We note that only three soil moisture readings were used: 2, 7, and 47.4. For each reading there were 9 values of \( Y, X_1, X_3, \) and \( X_4 \). Set up this problem as a completely randomized design with three treatments \( (X_2 = 2, 7, 47.4) \) and 9 plots per treatment.
(a) Analyze the effects of the three treatments on the amount of vitamin B₂ after adjusting for the effect of \( X_1 \).
(b) Analyze the effect of these three treatments after adjusting for \( X_1 \) and \( X_3 \).
(c) Can the independent variables be considered independent of the treatments in this example?

**Exercise 10.6.** Johnson and Tsao (7) analyzed the influence of sex, scholastic standing, individual order and grade on education development as measured by the **Iowa Tests of Education Development**. There were 3 of each of the last three variables and the 2 sexes, giving a total of 54 treatment combinations. An initial score and the mental age of each student was determined before the Development Tests were administered. Only one student was tested for each treatment combination. The final scores \( (Y) \), initial scores \( (X_1) \) and mental age \( (X_2) \) were as follows:
<table>
<thead>
<tr>
<th>Sex</th>
<th>Scholastic Standing</th>
<th>Ind. Order</th>
<th>Grade 10</th>
<th>Grade 11</th>
<th>Grade 12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Y</td>
<td>X&lt;sub&gt;1&lt;/sub&gt;</td>
<td>X&lt;sub&gt;2&lt;/sub&gt;</td>
</tr>
<tr>
<td>Male</td>
<td>Average</td>
<td>1</td>
<td>30</td>
<td>28</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>25</td>
<td>22</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>22</td>
<td>19</td>
<td>46</td>
</tr>
<tr>
<td>Male</td>
<td>Poor</td>
<td>1</td>
<td>26</td>
<td>22</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>17</td>
<td>14</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>14</td>
<td>14</td>
<td>29</td>
</tr>
<tr>
<td>Female</td>
<td>Average</td>
<td>1</td>
<td>18</td>
<td>18</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>17</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>12</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>Female</td>
<td>Poor</td>
<td>1</td>
<td>21</td>
<td>16</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>21</td>
<td>21</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>19</td>
<td>17</td>
<td>6</td>
</tr>
<tr>
<td>Female</td>
<td>Average</td>
<td>1</td>
<td>20</td>
<td>18</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>18</td>
<td>16</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>14</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>Female</td>
<td>Poor</td>
<td>1</td>
<td>14</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>12</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>9</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

SY = 1068; SX<sub>1</sub> = 944; SX<sub>2</sub> = 2379
SY<sup>2</sup> = 22730; SX<sub>1</sub><sup>2</sup> = 17926; SX<sub>2</sub><sup>2</sup> = 127369
SX<sub>1</sub>Y = 20116; SX<sub>2</sub>Y = 52005; SX<sub>1</sub>X<sub>2</sub> = 46227

(a) Set up an analysis of simple covariance on Y, using X<sub>2</sub> as the independent variate with the main effects of Sex, Scholastic Standing, Order, and Grade pulled out plus all two-factor interactions. Pool the three-factor and four-factor interactions as error.

(b) Repeat but with both independent variates.

(c) Evaluate this experiment.

**Exercise 10.7.** Covariance was used to reduce the experimental error in a randomized blocks experiment on Scuppernong grapes (9). 4 blocks and 5 magnesium treatments were used. 4 plants were used per plot. The yields were in terms of pounds per plant. Two independent variates were used to estimate yielding ability prior to treatment: X<sub>1</sub> = a score of 1 to 5 given by the investigators as to the vigor and size of each of the eight arms on each plant. X<sub>2</sub> = Diameter of each arm at a point 18 inches from the crown. For both X<sub>1</sub>'s, the individual arm measurements were cumulated...
for each plant. The sum of squares and cross products were as follows:

<table>
<thead>
<tr>
<th>D.F.</th>
<th>$S_y^2$</th>
<th>$S_{x_1}y$</th>
<th>$S_{x_2}y$</th>
<th>$S_{x_1}x_2$</th>
<th>$S_{x_1}^2$</th>
<th>$S_{x_2}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>19</td>
<td>1,001.90</td>
<td>306.25</td>
<td>629.55</td>
<td>308.64</td>
<td>154.52</td>
</tr>
<tr>
<td>Blocks</td>
<td>3</td>
<td>223.50</td>
<td>39.81</td>
<td>-85.95</td>
<td>9.07</td>
<td>20.64</td>
</tr>
<tr>
<td>Treatments</td>
<td>4</td>
<td>144.60</td>
<td>64.54</td>
<td>246.90</td>
<td>113.11</td>
<td>32.89</td>
</tr>
</tbody>
</table>

(a) Complete the analysis of covariance using only $X_1$. What was the efficiency of the covariance analysis?

(b) Repeat (a) using $X_2$ also.

(c) The treatment means per plant were as follows:

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y$</td>
<td>13.79</td>
<td>20.28</td>
<td>17.77</td>
<td>19.75</td>
<td>14.33</td>
<td>17.18</td>
</tr>
<tr>
<td>$X_1$</td>
<td>19.56</td>
<td>22.88</td>
<td>22.56</td>
<td>22.56</td>
<td>20.81</td>
<td>21.67</td>
</tr>
<tr>
<td>$X_2$</td>
<td>64.44</td>
<td>75.06</td>
<td>72.56</td>
<td>73.69</td>
<td>64.31</td>
<td>70.01</td>
</tr>
</tbody>
</table>

Derive the adjusted treatment means in (a) and (b) and the average standard error of the difference between two adjusted means.

**Exercise 10.8.** Covariance can also be used to make an analysis of variance when one or more of the plots are missing; this is only necessary when a double or higher restriction is made on the design, such as with a randomized blocks or a Latin Square design. The procedure is to set $Y = 0$ and $X = -1$ for the missing plot and $X = 0$ elsewhere. If there are several missing plots, multiple covariance must be used with $X_1 = -1$ for the first one, $X_2 = -1$ for the second, etc. Use the method of covariance to make the analysis of exercise 9.3.5. For a reference to these methods for other designs, see (9).

**Exercise 10.9.** For the application of covariance techniques to disproportionate frequency problems, see reference (10).

**Exercise 10.10.** The analysis of covariance is often used to determine if the same regression coefficient ($\beta$) applies to all treatments. Snedecor (5) presents an example of this type of analysis. Use the vitamin $B_2$ data (Exercise 10.5a) to compute a separate $\beta$ (for $X_1$) for each of the three treatments. Then set up the
following analysis, where $b_i$ is the regression coefficient of $Y$ on $X_i$ for each of the three treatments ($i = 1, 2, 3$).

<table>
<thead>
<tr>
<th>Source</th>
<th>D.F.</th>
<th>S.S.</th>
<th>M.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviations from average regression $(b)$</td>
<td>23</td>
<td>SSE*</td>
<td></td>
</tr>
<tr>
<td>Deviations from individual regressions</td>
<td>21</td>
<td>(SSE)*</td>
<td>$s^2n^2$</td>
</tr>
<tr>
<td>Differences among individual regressions</td>
<td>2</td>
<td>SSE* - (SSE)*</td>
<td>$[SSE^* - (SSE)^*]^2/2$</td>
</tr>
</tbody>
</table>

$$F = \left[ \frac{SSE^* - (SSE)^*}{(SSE)^*} \right]/2s^2n^2$$

$$(SSE)^* = S^2 - SST - \left[ b_1S_1y + b_2S_2y + b_3S_3y \right],$$

where $S$ applies to summation over treatment $i$.

**REFERENCES CITED**


**OTHER REFERENCES**


CHAPTER 11. COMPONENTS OF VARIANCE

11.1. Introduction

The regression models used in Chapters 8, 9, and 10 assumed that all variables were fixed except for a single random error term. Many experiments are designed so that several components are random instead of all except one being fixed. That is, the blocks in a randomized blocks design may be assumed to be randomly selected from a large population of blocks, so that the block to block part of the analysis of variance is also a random component. But the main use for an experimental model which postulates several random components is in sampling experiments; e.g., (i) a sample of soils to determine the basic sources of variability, such as plot-to-plot differences, sample-to-sample differences in the same plot and laboratory determination errors. (ii) A sample survey covering an entire region with a few counties selected from the large number of counties in the region, then a few areas selected from each sample county, and perhaps only one or two families from each selected area.

The regression model for this chapter can take on one of two forms: (i) every variable, except the general mean, is a random variable, or (ii) there is a mixture of random and fixed variables. Eisenhart (1) has presented the basic difference between (i) and the model used in chapters 8, 9, and 10 and has indicated the importance of (ii) without delving into the many theoretical difficulties involved in its use. Crump (2) presents the basic theory for (i) and includes an extensive bibliography of the use of components of variance. R. A. Fisher (3) indicated the additive properties of variances in the first edition of his Statistical Methods for Research Workers. Yates and Zacopanay (4) indicated the application of these methods to field sampling; among others, Cochran (5) extended them to enumerative surveys. Much of the recent development in the field of quantitative genetics has been built on a variance components model, such as in reference (6).

One of the authors of this book has presented the basic theory required for the mixed model (ii) in an article on the analysis of price data (7). The reader is cautioned that this last article has many drawbacks from an economic point of view,
but the difficulties encountered in the use of a mixed model are adequately outlined therein. This article drew much of its theory from articles by Daniels (8) and Satterthwaite (9). The mixed model is also required in the analysis of a series of experiments, such as those conducted over several years and at several places. Cochran and Cox (10) indicate some of the difficulties for this type of experimentation.

11.2. Analysis of Data with All Random Components, Except \( \mu \).

11.2.1. A Randomized Blocks Model. Let us assume that we have \( p \) treatments, each allocated to each of \( r \) blocks, and \( q \) samples taken from each of the \( pr \) plots, assuming that in each case the particular treatment, block, and sample is a random sample from an infinite (at least, very large) population. Hence there are \( prq \) samples. The model for the i-th sample from the j-th treatment on the k-th block is:

\[
y_{ijk} = \mu + \tau_i + \beta_j + (\tau \beta)_{ij} + \varepsilon_{ijk},
\]

where \( \tau_i, \beta_j, (\tau \beta)_{ij}, \) and \( \varepsilon_{ijk} \) are all assumed to be NID with means 0 and variances, \( \sigma^2_t, \sigma^2_b, \sigma^2_{tb}, \) and \( \sigma^2_e \), respectively. These variances are called variance components. Hence

\[
E(y_{ijk}) = \mu
\]

\[
\sigma^2(y_{ijk}) = \sigma^2_t + \sigma^2_b + \sigma^2_{tb} + \sigma^2_e
\]

The statistical problem is to estimate the variance components. The estimates will be designated as \( \hat{\sigma}^2 \), \( s^2 \) with the same subscripts. Obviously \( \mu \neq \bar{y} \).

Since we have more than one component of variation in this model, the method of least squares cannot be used in the estimation process. Instead we must utilize a more general estimation procedure, such as the method of maximum likelihood, which will be discussed in Part III. The method of least squares is the same as the method of maximum likelihood for the models used in the previous chapters of Part II.

In order to present this material in its practical connection with the analysis of variance problems discussed previously, we will assume that the data have been
analyzed by the analysis of variance, and then proceed to derive the expected values of the mean squares, using our variance components model. Hence we can derive unbiased estimates of the variance components, but this is no guarantee that the estimates are the most efficient which could be devised. A discussion of this matter will be deferred until Part III.

The analysis of variance pertaining to this model is:

Table: 

<table>
<thead>
<tr>
<th>Source</th>
<th>D.F.</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
<th>E(V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>r - 1</td>
<td>SSB</td>
<td>MSB = V₄</td>
<td>σₑ² + qσₜb² + pqσₑ² = σₑ²</td>
</tr>
<tr>
<td>Treatments</td>
<td>p - 1</td>
<td>SST</td>
<td>MST = V₃</td>
<td>σₑ² + qσₜb² + rpqσₑ² = σₑ²</td>
</tr>
<tr>
<td>T x B</td>
<td>(r - 1)(p - 1)</td>
<td>SS(TB)</td>
<td>MS(TB) = V₂</td>
<td>σₑ² + qσₜb² = σₑ²</td>
</tr>
<tr>
<td>Within Plots</td>
<td>(q - 1)rp</td>
<td>SSE</td>
<td>MSE = V₁</td>
<td>σₑ² = σₑ²</td>
</tr>
</tbody>
</table>

where we will use V to stand for a mean square.

SSB = \sum_j \frac{(B_j - G/r)^2}{pq} = \frac{\sum_i \left[ (B_j - pq\mu) - (G - rpq\mu) / r \right]^2}{pq}

\sum_j (B_j - pq\mu)^2 \frac{1}{pq} - \frac{(G - rpq\mu)^2}{rpq}

(B_j - pq\mu) = q \sum_i \tau_i + pq\beta_j + q \sum_i (\tau\beta)_{ij} + \sum_{ik} \sum E_{ijk}

(G - rpq\mu) = rq \sum_i \tau_i + pq \sum_j \beta_j + q \sum_i \sum_j (\tau\beta)_{ij} + \sum_{ik} \sum E_{ijk}

Hence

E(SSB) = \frac{r(pq^2\sigma^2_t + rp^2q^2\sigma^2_b + rpq^2\sigma^2_{tb} + pq\sigma^2_e)}{pq}

= (r - 1)pg \sigma^2_b + (r - 1)q \sigma^2_{tb} + (r - 1) \sigma^2_e
\[ E(V_4) = E(SSB)/(r - 1) = pq \sigma_b^2 + q \sigma_{tb}^2 + \sigma_e^2 = \sigma_2^2 \]

Since SST will be the same as SSB except for \( p \) and \( r \) and \( \sigma_t^2 \) and \( \sigma_b^2 \) interchanged, it is obvious that

\[
SST = \frac{\sum_i (T_i - r q \mu)^2}{r q} - \frac{(G - rpq \mu)^2}{rpq}
\]

\[ T_i - r q \mu = r q \tau_i + q \sum_j \beta_j + q \sum_j (T \beta)_{ij} + \sum_j \sum_k \varepsilon_{ijk} \]

\[ E(SST) = (p - 1)(r q \sigma_t^2 + q \sigma_{tb}^2 + \sigma_e^2) \]

\[ E(V_3) = r q \sigma_t^2 + q \sigma_{tb}^2 + \sigma_e^2 = \sigma_2^2 \]

Also we know that

\[
SS(TB) = \frac{\sum_i \sum_j [(TB)_{ij} - \frac{T_i}{r} - \frac{B_i}{p} + \frac{G}{rp}]^2}{q}
\]

\[
= \sum_i \sum_j \left[ (TB)_{ij} - q \mu - \frac{T_i - r q \mu}{r} - \frac{B_j - pq \mu}{p} + \frac{G - rpq \mu}{rp} \right]^2 \frac{1}{q}
\]

\[
= \sum_i \sum_j \frac{(TB)_{ij} - q \mu}{q} \left[ \frac{\sum_i (T_i - r q \mu)^2}{r q} - \frac{\sum_j (B_j - pq \mu)^2}{pq} \right] + \frac{(G - rpq \mu)^2}{rpq},
\]

where \((TB)_{ij} - q \mu = q \left[ \tau_i + \beta_j + (T \beta)_{ij} \right] + \sum_k \varepsilon_{ijk}\).

Hence

\[
E(SS(TB)) = rp \left[ q(\sigma_t^2 + \sigma_b^2 + \sigma_{tb}^2) + \sigma_e^2 \right] - p \left[ r q \sigma_t^2 + q \sigma_b^2 + q \sigma_{tb}^2 + \sigma_e^2 \right]
\]

\[
- \left[ q \sigma_t^2 + pq \sigma_b^2 + q \sigma_{tb}^2 + \sigma_e^2 \right] + \left[ (r q \sigma_t^2 + pq \sigma_b^2 + q \sigma_{tb}^2 + \sigma_e^2) \right]
\]

\[
= q(r p - r + 1) \sigma_{tb}^2 + (r p - r + 1) \sigma_e^2
\]

\[
= (r - 1)(p - 1)(q \sigma_{tb}^2 + \sigma_e^2).
\]

\[ E(V_2) = q \sigma_{tb}^2 + \sigma_e^2 = \sigma_2^2 \]
Finally we know that the total sum of squares is

\[ S_y^2 = S(Y - G/rpq)^2 = S \left[ \frac{Y - rpq \mu}{G - rpq \mu} \right]^2 - \frac{G - rpq \mu}{rpq} \]

The expected value of this total is

\[ rq(p - 1) \sigma^2_t + pq(r - 1) \sigma^2_b + q(rp - 1) \sigma^2_{tb} + (rpq - 1) \sigma^2_o. \]

Hence

\[ E(SSE) = E \left[ S_y^2 - SSE + SST - SS(TB) \right] = rp(q - 1) \sigma^2_o \]

\[ E(Y_1) = \sigma^2_o = \sigma^2_1 \]

The within residual also can be expressed as

\[ \epsilon_{ijk} = Y_{ijk} - (TB)_{ij}/q = \epsilon_{ijk} - \sum_k \epsilon_{ijk}/q \]

\[ E(\epsilon_{ijk}^2) = (q - 1) \sigma^2/o; \quad E(SSE) = rp(q - 1) \sigma^2_o. \]

Since

\[ \bar{Y} = \mu + \sum_i \frac{\tau_i}{r} + \sum_i \beta_i + \sum_i \sum_k (\tau \beta)_{ik} + \sum_i \sum_k \epsilon_{ijk} \]

\[ E(\bar{Y}) = \mu \quad \text{and} \quad \sigma^2(\bar{Y}) = \frac{\sigma^2_t}{p} + \frac{\sigma^2_b}{r} + \frac{\sigma^2_{tb}}{rp} + \frac{\sigma^2_o}{rq}. \]

Also if a particular treatment mean is of importance,

\[ \bar{T}_i = \frac{\sum_k Y_{ik}}{rq} = \mu + \tau_i + \frac{\sum_i \beta_i}{r} + \frac{\sum_i (\tau \beta)_{i1}}{r} + \frac{\sum_i \sum_k \epsilon_{ijk}}{rq} \]

\[ E(\bar{T}_i) = \mu + \tau_i \]

\[ \sigma^2(\bar{T}_i) = \frac{\sigma^2_t}{r} + \frac{\sigma^2_{tb}}{r} + \frac{\sigma^2_o}{rq}. \]

This assumes \( \tau_i \) is now a fixed variable.

The relative importance of various components can be assessed in \( \sigma^2(\bar{Y}) \) or \( \sigma^2(\bar{T}_i) \) and compared with the costs of obtaining the sample to enable the experimenter to better plan his future experiments. For example, if it costs \( \sigma_o \)
for each sample in a plot and \( a_p \) for each plot, then the total cost per treatment is

\[
C_1 = r q C_e + r C_p
\]

Suppose now that the total cost per treatment is fixed at \( C \). We are then led to minimize \( \sum\frac{\sigma^2(T)}{q} \) subject to the restriction \( C_1 = C \). In other words to select \( r \) and \( q \) so as to minimize

\[
-\frac{\sigma^2}{r q} + \lambda (r C_0) = 0; \quad -\frac{\sigma^2}{r q} (\sigma_e^2 + \sigma_{tb}^2 + \frac{\sigma^2}{q} + \lambda (r q C_e + C_p) = 0; \quad C_1 = C.
\]

The solutions are:

\[
q = \sqrt{\frac{C_p \sigma_e^2}{C_e (\sigma_b^2 + \sigma_{tb}^2)}}, \quad r = \frac{C}{q C_e + C_p}.
\]

Of course \( q \) and \( r \) must be integers, so the exact minimum won't be reached.

In order to make the estimates of the variance components \( \sigma_e^2, \sigma_{tb}^2, \sigma_t^2 \),

and \( \sigma_b^2 \) unbiased, we merely replace the components by their estimates

\( (\hat{e}_0^2, \hat{e}_{tb}^2, \hat{e}_t^2 \text{ and } \hat{e}_b^2 \text{ respectively}) \) in the \( E(V) \) column and equate respective rows in

the \( V \) and \( E(V) \) columns. Hence \( \sigma^2 = e^2 = V_1 \)

\[
\sigma_{tb}^2 = s_{tb}^2 = \frac{V_2 - V_1}{q}
\]

\[
\sigma_t^2 = s_t^2 = \frac{V_3 - V_2}{r q}
\]

\[
\sigma_b^2 = s_b^2 = \frac{V_4 - V_2}{p q}
\]

It is easy to show that all of the quantities

\[
\begin{bmatrix}
B_j - G/r \\
T_{ij} - G/p \\
Y_{ijk} - (TB)_{ij}/q \\
(TB)_{ij} - T_{ij}/r - B_j/p + G/rp
\end{bmatrix}
\]

are unbiased.
are independent of one another. Hence the various mean squares are independent.

Since any $V_i$ is a sum of squared linear functions of NID variates, it is independently distributed as $\chi^2(\sigma_i^2)$ with $f_i$ degrees of freedom, where $i$ stands for some particular line in the analysis of variance table with $f_i$ degrees of freedom in that line and $\sigma_i^2 = E(V_i)$. Also

$$\sigma^2(V_i) = \frac{2\sigma_i^4}{f_i}.$$

Since the mean squares are independent, the variance of an estimate of a variance component is simply a multiple of the sums of the variances of the mean squares used in estimating this variance component. That is, a variance component can be written in the form

$$s^2 = \sum_i a_i V_i / c$$

where $a_i = \pm 1$.

$$\sigma^2(s^2) = \sum_i \sigma^2(V_i) / c^2$$

$$= \frac{2c^2}{c^2} \left[ \sum V_i^2 / (f_i + 2) \right] \equiv s'^2.$$

For example,

$$s_b^2 = (V_1 - V_2) / pq$$

$$\sigma^2(s_b^2) = \left[ \sigma^2(V_1) + \sigma^2(V_2) \right] / (pq)^2$$

$$= \frac{2}{(pq)^2} \left[ \frac{V_1^2}{r + 1} + \frac{V_2^2}{(r - 1)(p - 1) + 2} \right] \equiv s'_b^2.$$

The problem of estimating confidence limits for $\sigma^2_b$ is still not completely solved. Some of the available procedures are indicated by Bross (11). These are:

1/This formula is slightly biased since $EV_i^2 = (f_i + 2) \sigma_i^4 / f_i$. Hence an unbiased estimate would be $2V_i^2 / (f_i + 2)$. Either of these two estimates is rather undesirable, since $s^2(V_i)$ is in terms of $V_i$ itself. Some attention needs to be paid to the use of the fourth moment of the observations in estimating $\sigma_i^4$. In some sampling problems, it may be possible to obtain several independent samples, determining $V_i$ from each, and then estimating $\sigma^2(V_i)$ from the sample-to-sample fluctuations.
(i) When \( r \) and \( p \) are large, the distribution of \( s_b^2 \) approaches normality. In this case the \((1 - 2\alpha)\) confidence limits are:

\[
s_b^2 - T_\alpha s_b < \sigma_b^2 < s_b^2 + T_\alpha s_b',
\]

where \( \Pr(T > T_\alpha) = \alpha \) and \( T \) is a normal deviate.

(ii) Satterthwaite (9) suggested that \( s^2 \) was approximately distributed as

\[
\frac{\chi^2}{\frac{f'}{f}} \sim \frac{s^2}{\sigma^2},
\]

with \( f' \) degrees of freedom, where

\[
f' = \left( \sum a_i^2 v_i \right)^2 / \sum (a_i^2 v_i^2 / f_i).
\]

In Chapter 14 it will be shown that the \((1 - 2\alpha)\) confidence limits for \( \sigma^2 \) are

\[
\frac{f' s^2}{\chi^2_{f'}} < \sigma^2 < \frac{f' s^2}{\chi^2(1 - \alpha)}
\]

where \( \Pr(\chi^2 > \chi^2_{f'}) = \Pr(\chi^2 < \chi^2(1 - \alpha)) = \alpha \). For

\[
\sigma_b^2, \quad f_b' = (v_4 - v_2)^2 / (\frac{v_4}{f_4} + \frac{v_2}{f_2}).
\]

Satterthwaite warned that when some of the \( a_i \)'s are negative, as in our case, caution must be exercised in the use of this \( \chi^2 \)-approximation.

(iii) If \( s^2 = (v_i - v_j) / c \) and \( \sigma^2 = E s^2 > 0 \),

\[
F_0 = \frac{v_i}{v_j} = F(1 + \frac{c}{\sigma^2_j} \frac{\sigma^2_j}{\sigma^2_j}
\]

where \( \sigma^2 \) is the variance component under consideration. In Chapter 14, it will be shown that the \((1 - 2\alpha)\) confidence limits for \((1 + c \sigma^2 / \sigma^2_j)\) are

\[
\frac{F_0}{F} < (1 + c \sigma^2 / \sigma^2_j) < \frac{F_0}{F(1 - \alpha)},
\]

where \( \Pr(F > F_\alpha) = \Pr(F < F(1 - \alpha)) = \alpha \), and \( F \) has \( f_i \) and \( f_j \)
degrees of freedom. If we replace \( \frac{\sigma^2}{c} \) by its estimate,

\[
\frac{V_j}{c} = \frac{s^2}{F_0 - 1},
\]

the confidence limits for \( \sigma^2 \) are

\[
\frac{F_0}{F_{\alpha}} - 1 \quad \frac{F_0}{F_0 - 1} \quad s^2 \leq \sigma^2 \leq \frac{F_0}{F_{(1-\alpha)}} - 1 \quad \frac{F_0}{F_0 - 1} \quad s^2.
\]

(iv) Using the fiducial probability concepts of R. A. Fisher (12), Bross (11) derived the following \((1 - 2\alpha)\) fiducial limits for \( \sigma^2 \):

\[
\begin{pmatrix}
\frac{F_0}{F_{\alpha}} - 1 \\
\frac{F_0}{F_{\alpha}} - 1
\end{pmatrix}
\leq \frac{F_0}{F_{(1-\alpha)}} - 1 \quad \frac{F_0}{F_{(1-\alpha)}} - 1
\]

where \( F_{\alpha}^{1} = F_{\alpha} \) for \((f_1\) and \(\infty)\) degrees of freedom.

In (iii) and (iv), \( F_{(1-\alpha)} = \frac{1}{F_{\alpha}} \), where \( F_{\alpha} = F_{\alpha} (f_1, f_2) \)

(see exercise 7.25). In all cases, if the lower limit is negative, it is replaced by zero. This will occur when \( F_0 \) is non-significant at the \( \alpha \) significance level \((F_0 < F_{\alpha})\).

**Example 11.1.** Let us consider the example presented by Grump (2) of a series of genetic experiments by J. W. Goven on the number of eggs laid by each of 12 females, from 25 races of *Drosophila melanogaster* on the fourth day of laying, the whole experiment being carried out 4 times \((r = 4, p = 25, q = 12)\). The analysis of variance was as follows:

\[
\begin{align*}
\text{degrees of freedom} & \quad \text{sum of squares} \\
\text{Female} & \quad 115.2 \quad 11.52 \\
\text{Race} & \quad 4.86 \quad 0.486 \\
\text{Error} & \quad 92.34 \quad 9.234 \\
\text{Total} & \quad 232.4 \quad 23.24 \\
\end{align*}
\]
Source of Variation | Degrees of Freedom | Mean Square | SS (MS) \\
--- | --- | --- | --- \\
Experiments (Blocks) | 3 | $V_4 = 46,659$ | $\sigma^2_4 = \sigma^2_e + 12 \sigma^2_{tb} + 300 \sigma^2_b$ \\
Races (Treatments) | 24 | $V_3 = 3,243$ | $\sigma^2_3 = \sigma^2_e + 12 \sigma^2_{tb} + 48 \sigma^2_t$ \\
T x B | 72 | $V_2 = 459$ | $\sigma^2_2 = \sigma^2_e + 12 \sigma^2_{tb}$ \\
Within Subclasses | 1100 | $V_1 = 231$ | $\sigma^2_1 = \sigma^2_e$ \\

$s^2_e = 231$ \\
$s^2_{tb} = \frac{459 - 231}{12} = 19$ \\
$s^2_t = \frac{3,243 - 459}{48} = 58$ \\
$s^2_b = \frac{46,659 - 459}{300} = 154$ \\
$s^2_e = 2(231)^2/1102 = 97$ \\
s^2_{tb} = \frac{2}{144} \left[ \frac{(231)^2}{1102} + \frac{(459)^2}{74} \right] = 40$ \\
s^2_t = \frac{2}{2304} \left[ \frac{(459)^2}{74} + \frac{(3243)^2}{26} \right] = 354$ \\
s^2_b = \frac{2}{90,000} \left[ \frac{(459)^2}{74} + \frac{(46,659)^2}{5} \right] = 9676$

The estimated variance of the general mean and a race mean would be

$s^2(T) = \frac{231}{r_p q} + \frac{19}{r_p} + \frac{58}{p} + \frac{154}{r}$

$s^2(T) = \frac{231}{r_q} + \frac{19}{r} + \frac{154}{r} = \frac{231}{r_q} + \frac{173}{r}$

The most important component appears to be the variation from experiment to experiment ($\sigma^2_b$). Hence in order to materially cut down the variance of a race mean, it is necessary to increase the number of experiments ($r$), since increasing $r$ decreases both parts of $s^2(T)$.

Since $\sigma^2_b$ is so important, let us consider the different methods of setting 90 per cent confidence limits for it.

<table>
<thead>
<tr>
<th>Method</th>
<th>5% Lower Limit</th>
<th>5% Upper Limit</th>
<th>10% Upper Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal (i)</td>
<td>0</td>
<td>316</td>
<td></td>
</tr>
<tr>
<td>$x^2$ (ii)</td>
<td>59</td>
<td>1344</td>
<td>806</td>
</tr>
<tr>
<td>$F$ (iii)</td>
<td>55</td>
<td>1331</td>
<td>801</td>
</tr>
<tr>
<td>Fiducial (iv)</td>
<td>58</td>
<td>1312</td>
<td>793</td>
</tr>
</tbody>
</table>
Some of the data needed for the above results were:

\[ s_b = 98.4 \quad f_b = 2.94 \quad F_0 = 101.65 \]

\[ \chi^2_{0.05} = 7.71 \quad \chi^2_{0.95} = 0.337 \quad \chi^2_{0.90} = 0.562 \]

\[ F_{0.05} = 2.74 \quad F_{0.95} = \frac{1}{8.57} \quad F_{0.90} = \frac{1}{5.16} \]

\[ F_{0.05} = 2.60 \quad F_{0.95} = \frac{1}{8.53} \quad F_{0.90} = \frac{1}{5.13} \]

We computed the 10 percent upper limit for the last three methods, assuming that the lower limit would be zero. Actually when all the probability is put on the upper tail the lower limit will be slightly negative for (iii) and (iv), but we have postulated that \( \sigma_b^2 \geq 0 \). When there are so few degrees of freedom for estimating a component such as \( \sigma_b^2 \), the confidence limits should be quite wide. The normal approximation is definitely unsatisfactory in this case. The other three estimates are remarkably close together. Bross (11) gives some examples where this is not the case. He criticizes the \( \chi^2 \)-method in one case of a non-significant \( F_0 \), for which methods (iii) and (iv) will give negative lower limits (assumed 0 since \( \sigma_b^2 \geq 0 \)), whereas the \( \chi^2 \) lower limit is greater than zero. It might be advisable, when \( F_0 \) is non-significant, to put all of the probability on the upper tail. This point should be considered from a theoretical standpoint.

11.2.2. Repeated Sub-sampling. A sampling technique usually used in sample surveys and in many sampling experiments in the physical and biological sciences is that of repeated sub-sampling, sometimes called nested sampling.\(^2\) We shall assume that there are four tiers in the universe: A, B in A, C in B, D in C. Obviously there may be more or less, but we shall illustrate the methods with four. Assume that the number of possible samples from each tier is large enough so that the sample represents only a small segment of each tier. Suppose that \( a \) A-units are selected, \( b \) B-units from each of the A-units, \( c \) C-units from each B-unit and \( d \) D-units from each C-unit. The model for a sample value is

\(^2\)See references (13) and (14).
\[ y_{ijklm} = \mu + \alpha_i + \beta_j + \gamma_{ijk} + \delta_{ijkm}, \]

where \( i = 1, 2, \ldots, a; \ j = 1, 2, \ldots, b; \ k = 1, 2, \ldots, s; \ m = 1, 2, \ldots, d. \) All of the effects, except \( \mu, \) are assumed NID \((0, \sigma^2_\theta),\) where \( \sigma^2_\theta = \sigma^2_a, \sigma^2_b, \sigma^2_c, \) or \( \sigma^2_d, \) respectively. Hence

\[ \bar{y} = \frac{\sum y_{ijklm}}{abed} \text{ and } \sigma^2(Y) = \frac{\sigma^2_a}{a} + \frac{\sigma^2_b}{ab} + \frac{\sigma^2_c}{abc} + \frac{\sigma^2_d}{abcd} \]

The analysis of variance is as follows:

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean Squares</th>
<th>E(MS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a - 1</td>
<td>( V_1 )</td>
<td>( \sigma^2_1 = \sigma^2_d + \sigma^2_c + cd \sigma^2_b + cbd \sigma^2_a )</td>
</tr>
<tr>
<td>B</td>
<td>a(b - 1)</td>
<td>( V_2 )</td>
<td>( \sigma^2_2 = \sigma^2_d + \sigma^2_c + cd \sigma^2_b )</td>
</tr>
<tr>
<td>C</td>
<td>ab(c - 1)</td>
<td>( V_3 )</td>
<td>( \sigma^2_3 = \sigma^2_d + \sigma^2_c )</td>
</tr>
<tr>
<td>D</td>
<td>abc(d - 1)</td>
<td>( V_4 )</td>
<td>( \sigma^2_4 = \sigma^2_d )</td>
</tr>
<tr>
<td>Total</td>
<td>abed - 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The sum of squares for \( A \) is computed like the block and treatment sums of squares in section 11.2.1. The remaining sums are simply within sums of squares:

\[
SSA = \frac{\sum A_i^2}{bcd} - \frac{G^2}{abcd}
\]

\[
SSB = \frac{\sum \sum B_{ij}^2}{cd} - \frac{\sum A_i^2}{bcd}
\]

\[\text{etc.,}\]

where \( A_i = \sum_j \sum_k \sum_m y_{ijklm}, \ B_{ij} = \sum_k \sum_m y_{ijklm}, \) etc. The expected values of the mean squares are computed as in section 11.2.1. All of the theory presented in that section pertaining to the estimates of the variance components and their variances and confidence limits carries over here, except that each variance component is estimated from its mean square and the one just below. For example

\[ s^2_a = \frac{V_1 - V_2}{bcd}. \]
Example 11.2. Unfortunately few sample surveys follow the sampling design outlined above, because the units are seldom of equal size so that the same number of sub-units are not selected from each. That is, we would not select b sub-units from each A-unit. Cochran (5) presents the analysis of a 1937 enumeration of commercial wheat fields in 6 districts of Great Britain. The sampling plan was to select a number of farms from each district, 1 or 2 fields per farm and 2 paths per field, each consisting of 6 bulked samples per path. In addition from the only farm in District III, 2 varieties were sampled in each of the three fields (2 paths per variety per field). The number of farms and fields per farm for each district were:

<table>
<thead>
<tr>
<th>District</th>
<th>No. Farms</th>
<th>No. Fields per Farm</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fields</td>
</tr>
<tr>
<td>I</td>
<td>2</td>
<td>2(2)</td>
<td>4</td>
</tr>
<tr>
<td>II</td>
<td>2</td>
<td>2, 1</td>
<td>3</td>
</tr>
<tr>
<td>III**</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>IV</td>
<td>9</td>
<td>2(2), 1(7)</td>
<td>11</td>
</tr>
<tr>
<td>V</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>VI</td>
<td>10</td>
<td>2(3), 1(7)</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td></td>
<td>36</td>
</tr>
</tbody>
</table>

*Numbers in parentheses refer to number of farms with this many fields.
**2 varieties were sampled in each field.

The analysis of variance, excluding the 3 degrees of freedom for varieties, in terms of cwt. per acre was:

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>( E(\text{MS}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Districts (A)</td>
<td>5</td>
<td>82.27</td>
<td>( \sigma_1^2 = 6.27^2 + 2.34\sigma_c^2 + 4.90\sigma_b^2 + 11.96\sigma_a^2 )</td>
</tr>
<tr>
<td>Farms in A(B)</td>
<td>19</td>
<td>52.73</td>
<td>( \sigma_2^2 = 2.00\sigma_c^2 + 2.58\sigma_b^2 )</td>
</tr>
<tr>
<td>Fields in B(C)</td>
<td>11</td>
<td>24.21</td>
<td>( \sigma_3^2 = \sigma_2^2 + 2.36\sigma_c^2 )</td>
</tr>
<tr>
<td>Paths in C(D)</td>
<td>39</td>
<td>6.60</td>
<td>( \sigma_4^2 = \sigma_2^2 )</td>
</tr>
</tbody>
</table>

The reader will note that the coefficients of \( \sigma_c^2 \) and \( \sigma_b^2 \) are not the same for all mean squares. See Ganguli (13) for the general details on this type problem.

The adventurous one might like to derive these expected values for, say \( \sigma_1^2 \).
\[ A_1 - 8 \mu = 8 \alpha_1 + 4 \sum_{j=1}^{2} \beta_{1j} + 2 \sum_{j,k=1}^{2} \gamma_{1jk} + \sum_{j,k,m=1}^{2} \delta_{1jkm} \]

\[ A_2 = 6 \mu = 6 \alpha_2 + (4 \beta_{21} + 2 \beta_{22}) + 2(\sigma_{211} + \sigma_{212} + \sigma_{221}) + \sum_{j,k} \sum_{2} \sigma_{2} \]

\[ A_3 - 12 \mu = 12 \alpha_3 + 12 \beta_{31} + 3 \sum_{k=1}^{3} \sigma_{31k} + \sum_{j,k} \sum_{3} \sigma_{3} \]

\[ G - n \mu = \sum_{i=1}^{6} (A_i - n_i \mu) \]

\[ S_{\text{SA}} = \sum_{i} \left( \frac{(A_i - n_i \mu)^2}{n_i} \right) \left( \frac{(G - n \mu)^2}{n} \right) \]

where \( n_i = \) number of paths in \( i \)th district and \( n = \sum n_i = 78 \).

\[ E \left( \frac{(A_1 - 8 \mu)^2}{8} \right) = 8 \sigma_a^2 + 4 \sigma_b^2 + 2 \sigma_c^2 + \sigma_d^2 \]

\[ E \left( \frac{(A_2 - 6 \mu)^2}{6} \right) = 6 \sigma_a^2 + \frac{20}{6} \sigma_b^2 + 2 \sigma_c^2 + \sigma_d^2 \]

\[ E \left( \frac{(A_3 - 12 \mu)^2}{12} \right) = 12 \sigma_a^2 + 12 \sigma_b^2 + 4 \sigma_c^2 + \sigma_d^2 \]

\[ E \left( \frac{(G - n \mu)^2}{n} \right) = \frac{(64 + 36 + 144 + \cdots)}{78} \sigma_a^2 + \frac{(32 + 20 + 144 + \cdots)}{78} \sigma_b^2 + \cdots \]

It should be noted that, for such unequal sub-sample sizes, the mean squares are no longer distributed as simply \( \chi^2 \sigma_i^2 / f_i \), but as a sum

\[ \lambda_1 \chi_1^2 + \lambda_2 \chi_2^2 + \cdots \], where the \( \lambda \)'s are functions of the variance
components and the number of observations. Hence the confidence limits presented
in section 11.2.1 cannot be used here. The theory applicable to this problem is
beyond this course.

Exercise 11.2.1. Prove that

\[ \sum_i \sum_j \left[ (TB)_{ij} - q \mu - \frac{T_i - rq \mu}{r} - \frac{B_j - pq \mu}{p} + \frac{G - rpq \mu}{rp} \right]^2 \]

\[ = \sum_i \sum_j \left[ (TB)_{ij} - q \mu \right]^2 - \frac{\sum_i (T_i - rq \mu)^2}{r} - \frac{\sum_i (B_j - pq \mu)^2}{p} \]

\[ + \frac{(G - rpq \mu)^2}{rp} \]

Exercise 11.2.2. (a) Prove that \((B_j - G/r)\) and \((T_i - G/p)\) are independent.
(b) Prove that each is independent of \( (TB)_{ij} - T_i/r - B_j/p + G/rp \).

Exercise 11.2.3. Show that \(EV_i^2 = \left( f_i + 2 \right) \frac{4}{f_1} \).

Exercise 11.2.4. (a) Given \( V_1 - V_j = c \chi^2 \sigma^2 / f_i \), where \( V_1 = \chi^2_i \sigma^2 / f_i \), \( V_j = \chi^2_j \sigma^2 / f_j \), \( c \sigma^2 = \sigma^2_i - \sigma^2_j \), and each \( \chi^2 \) has as degrees of freedom its denominator. Show by equating the variances of the two sides of this equation that

\[ f' = \frac{(c \sigma^2)^2}{\frac{\sigma^4_i}{f_i} + \frac{\sigma^4_j}{f_j}} \]

Also show that \( E(V_1 - V_j) = E(c \chi^2 \sigma^2 / f') \).

(b) How does this result connect with Satterthwaite's approximation for confidence limits on \( \sigma^2 \)?
Exercise 11.2.5. In example 11.1, suppose that \( C_e = 10, \ C_p = 120, \) and \( C = 960. \)
Solve for \( r \) and \( q \) to minimize \( s^2(T_i) \). What is \( s^2(T_i) \) using these values of \( r \) and \( q \)?
Compare this result with that obtained in the actual experiment \( (r = 4, \ q = 12) \).

Exercise 11.2.6. Show that the expected values of the mean squares given in
Section 11.2.2 are correct (Page 268).

Exercise 11.2.7. (a) Compute the estimates of the variance components for Cochran's
sample (example 11.2).

(b) In this example, these six districts constituted the population. Hence what
would you say regarding \( \sigma^2 \)?

(c) What would be the expected value and the variance of the general mean from
these data? Of the mean for District I?

Exercise 11.2.8. In Exercise 9.2.3 (page 183), suppose that the 13 markets were
chosen at random from a large population of markets.

(a) Set up the mathematical model for this sample and the expected values of the
mean squares in the analysis of variance. Show that the expected value of the
mean square for markets is \( \sigma^2 + k \sigma^2_m \), where \( \sigma^2_m \) is the variance
component for markets and

\[
\frac{1}{12} \left[ 69 - \frac{8^2 + 6^2 + \cdots + 9^2}{69} \right].
\]

(b) Estimate the variance components.

(c) What would be the estimated variance of the mean of a sample from \( r \) markets
with \( q \) sellers at each market?

(d) Suppose it costs $10.00 to visit a market and $1.00 to enumerate each seller,
what is the optimum number of markets and sellers if it is desired to obtain
100 schedules? How much will it cost to obtain this many schedules?
Exercise 11.2.9. In a 1940 Iowa AAA corn-accresage study, 2 sections were selected from each of 1617 townships. The analysis of variance of the corn acreage per section was

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between townships</td>
<td></td>
<td>6511.9</td>
</tr>
<tr>
<td>Within townships</td>
<td></td>
<td>1954.3</td>
</tr>
</tbody>
</table>

(a) Fill in the degrees of freedom and determine the expectations of the mean squares.

(b) Estimate the variance components.

(c) What is the variance of a sample mean if \( r \) townships and \( q \) sections per township were sampled?

(d) Determine confidence limits for the variance components.

Exercise 11.2.10. Rigney and Reed (15) studied some of the factors affecting the variability of estimates of various soil properties. They took samples from 20 fields (A), 2 sections (B) from each field, 2 samples (C) consisting of a composite of 20 borings from each section, and then 2 subsamples (D) from each sample. (\( a = 20, b = 2, c = 2, d = 2 \)). The analysis of variance for several of the properties is presented below:

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Calcium</td>
<td>Magnesium</td>
</tr>
<tr>
<td>A</td>
<td>19.19</td>
<td>0.1809</td>
</tr>
<tr>
<td>B in A</td>
<td>3.59</td>
<td>0.0545</td>
</tr>
<tr>
<td>C in B</td>
<td>0.30</td>
<td>0.0080</td>
</tr>
<tr>
<td>D in C</td>
<td>0.01</td>
<td>0.0005</td>
</tr>
</tbody>
</table>

(a) Fill in the degrees of freedom column.

(b) Set up the model for this sample and then the expectations of the mean squares in terms of the variance components.

(c) Estimate the variance components for one of the properties and the variance of the general mean for this sample. What is the variance of a mean of \( a \) fields, \( b \) sections/field, \( q \) samples/section, and \( d \) subsamples/sample?
Exercise 11.2.11. Marcuse (16) presents the following data on the moisture content of 2 cheeses from each of 3 different lots determined 2 times.

<table>
<thead>
<tr>
<th>Lot</th>
<th>Cheese 1</th>
<th>Cheese 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>39.02</td>
<td>38.96</td>
</tr>
<tr>
<td></td>
<td>38.79</td>
<td>39.01</td>
</tr>
<tr>
<td>II</td>
<td>35.74</td>
<td>35.58</td>
</tr>
<tr>
<td></td>
<td>35.41</td>
<td>35.52</td>
</tr>
<tr>
<td>III</td>
<td>37.02</td>
<td>35.70</td>
</tr>
<tr>
<td></td>
<td>36.00</td>
<td>36.04</td>
</tr>
</tbody>
</table>

(a) Set up the analysis of variance for these data with the expected values of the mean squares.

(b) Estimate the values of the variance components.

(c) Determine 90 per cent confidence limits for the "lots" component by the four methods outlined in section 11.2.1.

(e) Given that the costs are: 10 per lot, 3 per cheese per lot, and 1 per determination. Assume that 2 determinations per cheese are to be used and that the total cost is approximately 100, determine the number of lots and cheeses per lot to minimize the variance of the general mean. Could a lower variance be obtained by using other than 2 determinations per cheese?

Exercise 11.2.12. For a discussion of sampling from finite populations, see references (17) or (18).
11.3. Analysis of Data with Both Random and Fixed Effects.

11.3.1. Randomized Complete Blocks Designs. In most models of the type presented in section 9.3 and 11.2.1, the treatment \((\tau)\) effects are assumed to be fixed while the block \((\beta)\) effects are usually considered to be representative of a larger universe of environmental conditions than those of the particular experiment being analyzed. If both \((\tau)\) and \((\beta)\) are fixed, the interaction \((\tau\beta)\) effects are also assumed to be fixed. Hence in section 9.3, when no sampling error was specifically included, the interaction was assumed to be negligible and the interaction mean square was an estimate of \(\sigma^2_\beta\). That is, the correct model in section 9.3.1 is

\[
y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + e_{ij}
\]

Hence the expected value of \(s^2\) (in 9.3.3) is actually \(\sigma^2 = \sigma^2_\tau + \sigma^2_\beta\). Since there is no "within plots" mean square (as in 11.2.1), the experimenter cannot obtain separate estimates of \(\sigma^2_\tau\) and \(\sigma^2_\beta\). In this case, the experimenter has two courses of action open to him: (i) If \(\tau\) and \(\beta\) are fixed effects so that \((\tau\beta)_{ij}\) is also fixed, E(MST) will involve only \(\sigma^2_\tau\). Hence if \(s^2\) is to be used to test for treatment differences, \((\tau\beta)_{ij}\) must be assumed to be zero so that \(\sigma^2_\beta = 0\) and \(\sigma^2_\tau = s^2\). If there actually is a true interaction, the error term is biased upwards and \(F\) will be too small \([E(s^2) = \sigma^2_\tau + \text{constant}]\).

(ii) The second alternative is the subject of this section, namely that the \(\{\beta_j\}\) are assumed to be a sample from a larger universe. In this case

\[
\sigma^2 = \sigma^2_\tau + \sigma^2_\beta
\]

is the correct error term in 9.3.3, with \(\sigma^2_\beta\) a random component, because

\[
E(MST) = \sigma^2_\tau + \sigma^2_\beta + r \sigma^2_T
\]

Even for case (ii), the experimenter usually desires to know which of the two components of the error term \((\sigma^2_\tau\) and \(\sigma^2_\beta\)) is the more important, in order to plan his future experiments more efficiently. Hence he would set up the type of experiment of section 11.2.1 except that the \(p\) treatments \((\tau)\) would be the only
ones about which inferences were to be made (in other words, the treatment effects are fixed). The expressions for $T_i$, $B_j$, and $G$ given in section 11.2.1 also hold here. As then

$$E(\text{SSB}) = (r - 1)(pq \sigma_b^2 + q \sigma_{tb}^2 + \sigma_0^{-2})^2$$

$$E(\text{SS(TB)}) = (r - 1)(p - 1)(q \sigma_{tb}^2 + \sigma_0^{-2}),$$

where all effects are random. However

$$E(\text{SST}) = rq \sum T_i^2 + (p - 1)(q \sigma_{tb}^2 + \sigma_0^{-2})$$

$$E(V_3) = \frac{rq \sum T_i^2}{(p - 1)} + q \sigma_{tb}^2 + \sigma_0^{-2} \equiv rq \sigma_t^2 + \sigma_0^{-2},$$

where $\sigma_0^{-2}$ is estimated by $V_2$ as before, but $\sigma_t^{-2}$ is a fixed component and not a source of error.

In this case, $V_3$ is not distributed as $\chi^2 \sigma_0^{-2}/f_3$, because

$$E(T_i - rq \mu) = rq \tau_i \neq 0.$$

$V_3$ is distributed as a non-central $\chi^2$ ($EV_3 = \sigma_0^{-2} + \text{constant}$) with a variance of

$$\frac{4 \sigma_0^{-2}}{f_3} \left( \sigma_t^{-2} + rq \sigma_0^{-2} \right) - \frac{2 \sigma_0^{-2}}{f_3}.$$

\[2^1\] We know that $B_j - pq \mu = q \sum \tau_i + pq \beta_j + q \sum (T_i \beta)_{1j} + \sum \sum E_{ijk}$.

There is some question as to whether $\sum (T_i \beta)_{1j}$ sums to zero or not, when the $\tau_i$ are fixed (we know $\sum \tau_i = 0$). The above formula for $E(\text{SSB})$ assumes that the $(T_i \beta)_{1j}$ are purely random, even though the treatments are fixed, and do not sum to zero in any direction. We have used this traditional approach in this book, but feel that it has certain logical inconsistencies. If our results pertain to only these $n$ treatments, then regardless of the blocks used in the experiment, the same $n$ treatments will be included in every block. Hence it does not appear that there is a truly random interaction within a given block (summing over $i$). Perhaps some experimental sampling will be needed to settle the controversy. If $\sum (T_i \beta)_{1j} = 0$ when the treatments are fixed

$$E(\text{SSB}) = (r - 1)(pq \sigma_b^2 + \sigma_0^{-2}).$$

\[4^1\] See references (7) and (8).
It can be shown that an unbiased estimate of the variance of \( V_3 \), which consists of both random and fixed components, is

\[
s^2(V_3) = \frac{4}{f_3} v_3 v_2 - \frac{2 f_2 v_2^2}{(f_2 + 2) f_3}.
\]

In general if the random component consists of several sources of variation, which must be estimated from more than one mean square, we can write for the mean square (\( V \)) with both random and fixed components

\[
E(V) = \sigma^2_r + \lambda \sigma^2_s,
\]

where \( \sigma^2_r \) is the random and \( \sigma^2_s \) the fixed component. We can write

\[
\sigma^2_r = \sum a_j \sigma^2_j,
\]

where \( a_j = \pm 1 \) and \( \sigma^2_j \leq v_j \). That is, \( \sigma^2_r \) can be thought of as the sum and difference of several independent random variances, each estimated by a single mean square in the analysis of variance. Let \( V = V_r + V_s \), where \( \sigma^2_r \leq v_r \) and 

\[
\lambda \sigma^2_s \leq v_s.
\]

Since the \( V \)'s are all independent,

\[
v_r = \sum a_j v_j \quad \text{and} \quad E(VV_r) = EV_r EV = \sigma^2_r (\sigma^2_r + \lambda \sigma^2_s).
\]

Also since \( \sum f_j^2 = \sigma^2_r + 2 \sum a_j a_k \sigma^2_j \sigma^2_k \),

\[
\sigma^4_r = \sum \sigma^4_j + 2 \sum \sum a_j a_k \sigma^2_j \sigma^2_k
\]

\[
= \sum \frac{f_j}{f_j + 2} v_j^2 + 2 \sum \sum a_j a_k v_j v_k.
\]

Collecting terms we find that an unbiased estimate of the variance of a mean square \( V_r \), consisting of both random and fixed components, is

\[
\frac{4}{f} v V_r - \frac{2}{f} \left[ \sum \frac{f_j}{f_j + 2} v_j^2 + 2 \sum \sum a_j a_k v_j v_k \right],
\]
where \( V_r = \sum a_j V_j \) and \( \text{degrees of freedom for } V \).

**Example 11.2.** In a bee experiment conducted by E. C. Oertel of Baton Rouge, Louisiana, the honey was collected from 22 colonies randomly assigned to 4 rows of 5 colonies each for each of 5 years. We assume that the 5 years (1942-1946) were representative of a larger population of years and the 5 colonies for each row were also a random sample of possible colonies but that the 4 row effects were fixed. The experimenter wanted to determine if there were row-to-row differences and if the year-to-year fluctuations were significantly greater than the sampling fluctuation from colony to colony. The model for this experiment is

\[
Y_{ijk} = \mu + \tau_i + \beta_k + (\tau \beta)_{ik} + \gamma_{ij} + \epsilon_{ijk},
\]

where \( \tau_i \) is the effect of the \( i \)-th row, \( \beta_k \) the random component for the \( k \)-th year, \( (\tau \beta)_{ik} \) the random row-treatment interaction, \( \gamma_{ij} \) the random component for the \( j \)-th colony in the \( i \)-th row and \( \epsilon_{ijk} \) the random colony-year interaction.

The analysis of variance of the yields in pounds of honey was

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>E(MS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years (B)</td>
<td>4</td>
<td>107,297 = V_b</td>
<td>( \sigma^2_e + 8 \sigma^2_{tb} + 32 \sigma^2_b )</td>
</tr>
<tr>
<td>Rows (T)</td>
<td>3</td>
<td>15,593 = V_t</td>
<td>( \sigma^2_e + 5 \sigma^2_{tb} + 8 \sigma^2_{tb} + 40 \sigma^2_t )</td>
</tr>
<tr>
<td>T x B</td>
<td>12</td>
<td>3,019 = V_{tb}</td>
<td>( \sigma^2_e + 8 \sigma^2_{tb} )</td>
</tr>
<tr>
<td>Colonies (in Rows)</td>
<td>28</td>
<td>2,529 = V_c</td>
<td>( \sigma^2_e + 5 \sigma^2_c )</td>
</tr>
<tr>
<td>C x B</td>
<td>112</td>
<td>2,589 = V_e</td>
<td>( \sigma^2_e )</td>
</tr>
</tbody>
</table>

The following conclusions can be made:

(i) Since \( V_c \) is actually less than \( V_e \), we conclude that \( \sigma^2_c \) is non-significant.

(ii) \( \sigma^2_{tb} \) is also not significantly greater than zero, using \( F = 3019/2589 = 1.17 \) with 12 and 112 degrees of freedom.

(iii) \( \sigma^2_b \) is unquestionably greater than zero, as shown by \( F = 107,297/3,019 = 35.5 \). Hence it is very important to secure as many year's data as possible.
if some over-all production per colony is required.

(iv) In this case the test for row effects is quite simple, because $\sigma^2_c$ can be disregarded. The appropriate error term is $V_{tb}$; hence, $F = 15,593/3,019 = 5.16$, with 3 and 12 degrees of freedom, a significant value. If we could not disregard $\sigma^2_c$, the appropriate error term might be estimated as

$$V_r = V_{tb} + V_c - V_o$$

(See page 277). The main difficulty with the use of this error term is that it is not distributed as $\chi^2 \sigma^2_r/f_r$, where

$$\sigma^2_r = \sigma^2_c + 5 \sigma^2_c + 8 \sigma^2_{tb}.$$ It is suggested that Satterthwaite's approximation (9), mentioned in section 11.2.1, be used to estimate $f_r$, and then assume that $F = V_t/V_r$ is approximately distributed as $F$ with 3 and $f_r$ degrees of freedom. In our case, $V_r = 3,019 + 2,529 - 2,589 = 2,959$

$$f_r = \frac{V_r^2}{\frac{V_{tb}^2}{f_{tb}} + \frac{V_c^2}{f_c} + \frac{V_o^2}{f_o}} = 8.4.$$

$$F' = 15,593/2,959 = 5.27$$, a significant value (3 and 8.4 degrees of freedom).

11.3.2. Split-plot Designs. A special type of incomplete blocks design is the split-plot design, in which one set of $p$ treatments ($T$) is arranged in a randomized complete blocks design with $r$ blocks ($B$) and a second set of $q$ treatments ($A$) is assigned at random to one of $q$ sub-plots in each of the $pr$ whole plots. The mathematical model for this experiment is

$$Y_{ijk} = \mu + T_i + \beta_j + (T\beta)_{ij} + \alpha_k + (T\alpha)_{ik} + \varepsilon_{ijk},$$

where $T$, $\alpha$ and $(T\alpha)$ are fixed treatment effects $(T\beta)_{ij}$ and $\varepsilon_{ijk}$ are random sources of error, and $\beta_j$ is a random block effect. The analysis of

---

5/ See reference (7) for a more complete discussion of this point.

6/ In this case we assume that there is no real $(\alpha \beta)$ interaction, but that the only deviation in the sub-plot is a single random error, $\varepsilon_{ijk}$. 
The variance for this model is:

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
<th>E(Mean Squares)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks (B)</td>
<td>r - 1</td>
<td>SSB</td>
<td>V_b</td>
<td>σ_e^2 + q σ_{tb}^2 + pq σ_b^2</td>
</tr>
<tr>
<td>T</td>
<td>p - 1</td>
<td>SST</td>
<td>V_t</td>
<td>σ_e^2 + q σ_{tb}^2 + rq σ_t^2</td>
</tr>
<tr>
<td>T x B</td>
<td>(r - 1)(p - 1)</td>
<td>SS(TB)</td>
<td>V_{tb}</td>
<td>σ_e^2 + q σ_{tb}^2</td>
</tr>
<tr>
<td>A</td>
<td>(q - 1)</td>
<td>SSA</td>
<td>V_a</td>
<td>σ_e^2 + rp σ_a^2</td>
</tr>
<tr>
<td>T x A</td>
<td>(p - 1)(q - 1)</td>
<td>SS(TA)</td>
<td>V_{ta}</td>
<td>σ_e^2 + r σ_{ta}^2</td>
</tr>
<tr>
<td>Sub-plot Error</td>
<td>p(r - 1)(q - 1)</td>
<td>SSE</td>
<td>V_e</td>
<td>σ_e^2</td>
</tr>
</tbody>
</table>

The first five sums of squares are computed in the usual manner for main effects and interaction (see section 9.5), and the sub-plot error (SSE) by subtraction. As mentioned above, σ_e^2, σ_{tb}^2, and σ_b^2 are true variances, while σ_t^2, σ_a^2, and σ_{ta}^2 are fixed components. Again there is disagreement as to whether (q σ_{tb}^2) should be included in the expectation of the Blocks mean square.

In making tests of significance of the differences among fixed effects, SSE is the error sum of squares for A and TA and SS(TB) for T. The problem of determining confidence limits for differences between various treatment means is slightly complicated. We know that

\[(TA)_{ik} = r \mu + r \bar{T}_i + \sum_j \beta_j + \sum_j (T \beta)_{ij} + r \alpha_k + r(T \alpha)_{ik} + \sum_j \varepsilon_{ijk} \]

We have the following types of means to compare:

(1) Two T treatments,

\[ d_t = \bar{T}_i - \bar{T}_{i'} = (\bar{T}_i - \bar{T}_{i'}) + \sum_j \left[ (T \beta)_{ij} - (T \beta)_{i'j} \right] / r + \sum_j \sum_k (\varepsilon_{ijk} - \varepsilon_{i'jk}) / rq \]

\[ E(d_t) = (\bar{T}_i - \bar{T}_{i'}) ; \quad \sigma^2(d_t) = \frac{2}{rq} \left[ q \sigma_{tb}^2 + \sigma_e^2 \right] + 2V_{tb} / rq \]
(ii) Two A treatments.

\[ d_a = \bar{a}_k - \bar{a}_{k'} = (\alpha_k - \alpha_{k'}) + \sum_i \sum_j (\epsilon_{ijk} - \epsilon_{ijk'})/r^p \]

\[ E(d_a) = \alpha_k - \alpha_{k'} \quad \sigma^2(d_a) = 2 \sigma^2_e/r^p \approx 2V_e/r^p. \]

(iii) Two T treatments with the same A treatment.

\[ d_{tk} = (\bar{T}A)^i_{ik} - (\bar{T}A)^i{ik'} = (\bar{T}_i - \bar{T}_{i'}) + \sum_j \left[ (\bar{T}\beta)^i_{ij} - (\bar{T}\beta)^i{ij'} \right]/r \]

\[ + \left[ (\bar{T}\alpha)^i_{ik} - (\bar{T}\alpha)^i{ik'} \right] + \sum_j (\epsilon_{ijk} - \epsilon_{ijk'})/r \]

\[ E(d_{tk}) = \bar{T}_i - \bar{T}_{i'} + \left[ (\bar{T}\alpha)^i_{ik} - (\bar{T}\alpha)^i{ik'} \right] \]

\[ \sigma^2(d_{tk}) = 2(\sigma^2_{tb} + \sigma^2_e)/r \approx 2 \left[ V_{tb} + (q - 1)V_e \right]/r^q. \]

(iv) Two A treatments with the same T treatment.

\[ d_{ia} = (\bar{T}A)^i_{ik} - (\bar{T}A)^i{ik'} = (\alpha_k - \alpha_{k'}) + \left[ (\bar{T}\alpha)^i_{ik} - (\bar{T}\alpha)^i{ik'} \right] + \sum_j (\epsilon_{ijk} - \epsilon_{ijk'})/r \]

\[ E(d_{ia}) = \alpha_k - \alpha_{k'} + \left[ (\bar{T}\alpha)^i_{ik} - (\bar{T}\alpha)^i{ik'} \right] \]

\[ \sigma^2(d_{ia}) = 2 \sigma^2_e/r. \]

(v) Two different A treatments and two different T treatments.

\[ d_{ta} = (\bar{T}A)^i_{ik} - (\bar{T}A)^i{ik'} = (\bar{T}_i - \bar{T}_{i'}) + \sum_j \left[ (\bar{T}\beta)^i_{ij} - (\bar{T}\beta)^i{ij'} \right]/r \]

\[ + \left[ (\bar{T}\alpha)^i_{ik} - (\bar{T}\alpha)^i{ik'} \right] + \sum_j (\epsilon_{ijk} - \epsilon_{ijk'})/r \]

\[ E(d_{ta}) = \bar{T}_i - \bar{T}_{i'} + (\alpha_k - \alpha_{k'}) + \left[ (\bar{T}\alpha)^i_{ik} - (\bar{T}\alpha)^i{ik'} \right] \]

\[ \sigma^2(d_{ta}) = 2(\sigma^2_{tb} + \sigma^2_e)/r \approx 2 \left[ V_{tb} + (q - 1)V_e \right]/r^q. \]
A more complete description of the split-plot design can be found in references (10) and (19). This design enables the experimenter to obtain quite accurate information on the A treatments and on the interaction between T and A at the expense of less accurate information on the T treatments as compared with a factorial design (section 9.5), because the estimate of the whole-plot error is based on fewer degrees of freedom than that of the sub-plot error. Also with this design, the experimenter can allocate the T treatments to rather large plots and reserve the small sub-plots for the A treatment. If one set of treatments is cultivation methods and a second set is fertilizer treatments, it would be advantageous to put the former on the larger whole plots because of the difficulty of handling machinery on small plots.

Example 11.6. Wilm (20) presents the following analysis of variance for data on soil moisture deficits (in inches of water) as affected by timber cutting with 4 blocks, 5 T-treatments (amount of cutting) and 3 years as sub-plot treatments, (the data are presented on the next page).

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>E(MS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks (B)</td>
<td>3</td>
<td>1.4832</td>
<td>( \sigma^2_c + 3 \sigma^2_{cb} + 15 \sigma^2_b )</td>
</tr>
<tr>
<td>Treatments (T)</td>
<td>4</td>
<td>1.3333</td>
<td>( \sigma^2_c + 3 \sigma^2_{cb} + 12 \sigma^2_t )</td>
</tr>
<tr>
<td>T x B</td>
<td>12</td>
<td>0.3909</td>
<td>( \sigma^2_c + 3 \sigma^2_{cb} )</td>
</tr>
<tr>
<td>Years (A)</td>
<td>2</td>
<td>6.5418</td>
<td>( \sigma^2_e + 20 \sigma^2_a )</td>
</tr>
<tr>
<td>T x A</td>
<td>8</td>
<td>0.2554</td>
<td>( \sigma^2_e + 4 \sigma^2_{ta} )</td>
</tr>
<tr>
<td>Sub-plot Error</td>
<td>30</td>
<td>0.1101</td>
<td>( \sigma^2_e )</td>
</tr>
</tbody>
</table>

As indicated above, \( \sigma^2_e \), \( \sigma^2_{cb} \) and \( \sigma^2_b \) are assumed to be true variances and \( \sigma^2_{ta} \), \( \sigma^2_a \) and \( \sigma^2_t \) fixed components. If the experimenter could assume that these 3 years were representative of a larger universe of time, \( \sigma^2_a \) and \( \sigma^2_{ta} \) would be random components and 4 \( \sigma^2_{ta} \) would appear in the expectation of the "years" and "treatments" mean squares. Assuming the year-effects are fixed, the T x A and A effects have 0.1101 as their error term and the T effects have 0.3909 as its error term.
Soil Moisture Deficits as Affected by Timber Cutting (All data in inches of water).

<table>
<thead>
<tr>
<th>*Treatment and Year</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Treatment sums</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unout (11,900)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1941</td>
<td>2.40</td>
<td>0.98</td>
<td>1.38</td>
<td>1.37</td>
<td>6.13</td>
</tr>
<tr>
<td>1942</td>
<td>3.32</td>
<td>1.91</td>
<td>2.36</td>
<td>1.62</td>
<td>9.21</td>
</tr>
<tr>
<td>1943</td>
<td>2.59</td>
<td>1.44</td>
<td>1.66</td>
<td>1.75</td>
<td>7.44</td>
</tr>
<tr>
<td>Sum</td>
<td>8.31</td>
<td>4.33</td>
<td>5.40</td>
<td>4.74</td>
<td>22.78</td>
</tr>
<tr>
<td>6,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1941</td>
<td>1.76</td>
<td>1.65</td>
<td>1.69</td>
<td>1.11</td>
<td>6.21</td>
</tr>
<tr>
<td>1942</td>
<td>2.78</td>
<td>2.07</td>
<td>2.98</td>
<td>2.50</td>
<td>10.33</td>
</tr>
<tr>
<td>1943</td>
<td>2.27</td>
<td>2.28</td>
<td>2.16</td>
<td>2.06</td>
<td>8.77</td>
</tr>
<tr>
<td>Sum</td>
<td>6.81</td>
<td>6.00</td>
<td>6.83</td>
<td>5.67</td>
<td>25.31</td>
</tr>
<tr>
<td>4,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1941</td>
<td>1.43</td>
<td>1.30</td>
<td>0.18</td>
<td>1.66</td>
<td>4.57</td>
</tr>
<tr>
<td>1942</td>
<td>2.51</td>
<td>1.48</td>
<td>1.83</td>
<td>2.36</td>
<td>8.18</td>
</tr>
<tr>
<td>1943</td>
<td>1.54</td>
<td>1.46</td>
<td>0.16</td>
<td>1.84</td>
<td>5.00</td>
</tr>
<tr>
<td>Sum</td>
<td>5.48</td>
<td>4.24</td>
<td>2.17</td>
<td>5.86</td>
<td>17.75</td>
</tr>
<tr>
<td>2,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1941</td>
<td>1.24</td>
<td>0.70</td>
<td>0.69</td>
<td>0.82</td>
<td>3.45</td>
</tr>
<tr>
<td>1942</td>
<td>3.29</td>
<td>2.00</td>
<td>1.38</td>
<td>1.98</td>
<td>8.65</td>
</tr>
<tr>
<td>1943</td>
<td>2.67</td>
<td>1.44</td>
<td>1.75</td>
<td>1.56</td>
<td>7.42</td>
</tr>
<tr>
<td>Sum</td>
<td>7.20</td>
<td>4.14</td>
<td>3.82</td>
<td>4.36</td>
<td>19.52</td>
</tr>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1941</td>
<td>0.79</td>
<td>0.21</td>
<td>0.01</td>
<td>0.16</td>
<td>1.17</td>
</tr>
<tr>
<td>1942</td>
<td>1.70</td>
<td>1.44</td>
<td>2.65</td>
<td>2.15</td>
<td>7.94</td>
</tr>
<tr>
<td>1943</td>
<td>1.62</td>
<td>1.26</td>
<td>1.36</td>
<td>1.87</td>
<td>6.11</td>
</tr>
<tr>
<td>Sum</td>
<td>4.11</td>
<td>2.91</td>
<td>4.02</td>
<td>4.18</td>
<td>15.22</td>
</tr>
</tbody>
</table>

**Block Sums**

<table>
<thead>
<tr>
<th>Year</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Treatment sums</th>
</tr>
</thead>
<tbody>
<tr>
<td>1941</td>
<td>7.62</td>
<td>4.84</td>
<td>3.95</td>
<td>5.12</td>
<td>21.53</td>
</tr>
<tr>
<td>1942</td>
<td>13.60</td>
<td>8.90</td>
<td>11.20</td>
<td>10.61</td>
<td>44.31</td>
</tr>
<tr>
<td>1943</td>
<td>10.69</td>
<td>7.88</td>
<td>7.09</td>
<td>9.08</td>
<td>34.74</td>
</tr>
<tr>
<td>Sum</td>
<td>31.91</td>
<td>21.62</td>
<td>22.24</td>
<td>24.81</td>
<td>100.58</td>
</tr>
</tbody>
</table>

*Expressed as volume in board-feet of trees larger than 9.6 inches in diameter, which were left in the forest after treatment.*
11.3.3. Simple Lattice Designs. In section 9.7.2 (page 225), we indicated that most analyses of lattice designs make use of "inter-block information" to increase the efficiency of the estimates of treatment effects. The basic change in the model is that the $\beta$'s are assumed to be random effects with the same variance, $\sigma^2 \beta$.

Since the blocks are arranged in complete replications, a constant $\gamma$ is introduced into the model for fixed replication effects, with random $\beta$ effects within replications. As before we assume that we have $p = k^2$ treatments in $k$ blocks of $k$ treatments per block. Since the simple lattice has $2$ replications, we have $2k$ blocks and $2k^2$ plots.

The first replication is designated as $X$ and the second as $Y$ with individual yields $X_{ij}$ and $Y_{ij}$, where $(i, j) = 1, 2, \ldots, k$. The mathematical model for $X_{ij}$ and $Y_{ij}$ is

$$
X_{ij} = \mu + \gamma x + \beta x_i + \tau_{ij} + e_{ij} \\
Y_{ij} = \mu + \gamma y + \beta y_j + \tau_{ij} + e'_{ij},
$$

where $\beta$, $e$, and $e'$ are independent random effects and the others are fixed. $\{\beta\}$ are NID $(0, \sigma^2 \beta)$ and $\{\gamma, e, e'\}$ are NID $(0, \sigma^2 \gamma)$. Hence

$$
G = 2k^2 \mu + k(\sum_i \beta x_i + \sum_j \beta y_j) + \sum \sum (e'_{ij} - e_{ij})
$$

$$
X_{..} = k^2(\mu + \gamma x) + k \sum_i \beta x_i + \sum \sum e_{ij} \\
Y_{..} = k^2(\mu + \gamma y) + k \sum_j \beta y_j + \sum \sum e'_{ij}
$$

$$
X_{i.} = k(\mu + \gamma x) + k \beta x_i + \sum_j \tau_{ij} + \sum e_{ij} \\
Y_{i.} = k(\mu + \gamma y) + \sum_j \beta y_j + \sum_j \tau_{ij} + \sum e'_{ij}
$$

$$
D_{..} = X_{i.} - Y_{i.} = k(\gamma x - \gamma y) + k \beta x_i - \sum_j \beta y_j + \sum_j e_{ij} - e'_{ij} \\
D_{ij} = k(\gamma x - \gamma y) + k(\beta x_i - \beta y_j) + \sum_i \sum_j (e_{ij} - e'_{ij}),
$$
where \( D_{ij} \) is a linear function of the block effects adjusted for treatment effects. Similarly we can compute \( D_{ij} \) using \( X_{ij} \) and \( Y_{ij} \). These \( (2k) D's \) could be used to solve for the \( \beta \)'s, adjusted for the \( \zeta \)'s.

Obviously we can use \( G, X_{..} \) and \( Y_{..} \) to estimate \( \mu, \sigma_x \) and \( \sigma_y \). The replication sum of squares in the analysis of variance (with one degree of freedom) is

\[
\frac{X_{..}^2 + Y_{..}^2}{k^2} - \frac{G^2}{2k^2} = \frac{(X_{..} - Y_{..})^2}{2k^2}
\]

with an expected value of \( k^2 (\frac{\sigma_x^2}{2} + \frac{\sigma_y^2}{2}) + k (\sigma_b^2 + \sigma_e^2) \).

A block sum of squares, adjusted for treatments, with \( 2(k - 1) \) degrees of freedom, can be computed from

\[
SSB(\text{adj.}) = \frac{\sum D_{i..}^2 + \sum D_{.j}^2}{2k} - \frac{2D_{..}^2}{2k^2}.
\]

\[
E\left[SSB(\text{adj.})\right] = 2(k - 1)(\frac{k}{2} \sigma_b^2 + \sigma_e^2).
\]

Hence if we designate \( V_b = SSB(\text{adj.})/2(k - 1) \),

\[
E(V_b) = \frac{k}{2} \sigma_b^2 + \sigma_e^2.
\]

It can be shown that

\[
SSE = Sx^2 + Sy^2 - SST(\text{unadj.}) - SSB(\text{adj.}),
\]

with \( 2(k^2 - 1) - (k^2 - 1) - 2(k - 1) = (k - 1)^2 \) degrees of freedom. Also

\[
E(V_o) = E(\text{MSE}) = \sigma_e^2.
\]

So that

\[
\sigma_e^2 \leq V_o \quad \text{and} \quad \sigma_b^2 \leq 2(V_b - V_o)/k.
\]

In order to estimate a treatment effect, \( t_{ij} \), we shall write \( \hat{t} \)

\( \hat{t} \)

This method is suggested by the similarity between this design and a factorial with \( k \) levels of two treatments. The effect of the \( i \)th level of one and \( j \)th level of a second is generally written in this form. Let \( \hat{1} \) stand for the mean effects (first parentheses) and \( \hat{0} \) for the interaction (second parentheses).
\[ t_{ij} = \bar{t}_{i1}^* + \bar{t}_{j1}^* + (t_{ij} - \bar{t}_{i1}^* - \bar{t}_{j1}^*) \]
\[ = (\bar{t}_{i1}^* + \bar{t}_{j1}^* - 2m) + (t_{ij} - \bar{t}_{i1}^* - \bar{t}_{j1}^* + m) \equiv t + v. \]

In section 9.7.2, we used \( \bar{Y}_{i1} \) as an estimate of \( \bar{t}_{i1}^* \), and \( \bar{Y}_{j1} \) as an estimate of \( \bar{t}_{j1}^* \), since these two estimates were free of block effects. However now that we have a means of handling the block effects, it seems reasonable to utilize this "inter-block information."

It turns out that the part in the second parentheses \((v)\) is independent of block effects, even when we use all of the data to estimate \( \bar{t}_{i1}^* \) and \( \bar{t}_{j1}^* \). That is if we use all of the data,

\[ \bar{t}_{i1}^* = \frac{Y_{i1} + Y_{i1}^*}{2k} = \mu + \bar{t}_{i1} + \frac{\beta_x Y_{i1}^*}{2} + \frac{\beta_y Y_{i1}}{2} + \frac{\bar{e}_{i1}^* + \bar{e}_{i1}}{2} \]

\[ \bar{t}_{j1} = \frac{Y_{j1} + Y_{j1}^*}{2k} = \mu + \bar{t}_{j1} + \frac{\beta_x Y_{j1}^*}{2} + \frac{\beta_y Y_{j1}}{2} + \frac{\bar{e}_{i1} + \bar{e}_{i1}^*}{2} \]

\[ t_{ij} = \frac{X_{ij} + X_{ij}^*}{2} = \mu + \bar{t}_{ij} + \frac{\beta_x Y_{ij}^*}{2} + \frac{\beta_y Y_{ij}}{2} + \frac{\bar{e}_{ij}^* + \bar{e}_{ij}}{2} \]

\[ m = \frac{G}{2k^2} = \mu + \frac{\beta_x Y_{..}^*}{2} + \frac{\beta_y Y_{..}}{2} + \frac{\bar{e}_{..} + \bar{e}_{..}^*}{2} \]

where \( \bar{e}_{i1} = \sum_j E_{ij}/k, \bar{e}_{i1} = \sum \sum E_{ij}/k^2 \), etc. Hence

\[ p \equiv (t_{ij} - \bar{t}_{i1}^* - \bar{t}_{j1}^* + m) = \bar{t}_{ij} - \bar{t}_{i1}^* - \bar{t}_{j1}^* + \bar{e}_{ij} + \bar{e}_{ij}^* + \bar{e}_{ij} + \bar{e}_{ij}^* \]

where

\[ \bar{e}_{ij} = \frac{1}{2} \left[ \bar{e}_{ij} + \bar{e}_{ij}^* - (\bar{e}_{i1} + \bar{e}_{i1}^* + \bar{e}_{j1} + \bar{e}_{j1}^*) + \right. \]

\[ \left. + (\bar{e}_{ij} + \bar{e}_{ij}^*) \right]^1. \]
However if we use all of the data to estimate \( t = (t_1^* + t_2^*) - 2m \), block effects will be involved. Hence it is proposed that we obtain two different estimates of \( t_1^* \), one free of block effects \( (t_1) \) and one involving block effects \( (t_2) \), and form a weighted average of the two.

\[
t = \frac{W_1 t_1 + W_2 t_2}{W_1 + W_2},
\]

where \( W_i \) will be inversely proportional to the variance of \( t_i^* \). Then we will have

\[
t_1 = \frac{Y_i + X_i}{k} - \frac{G}{k^2} = \bar{t}_j + \bar{t}_i + \varepsilon_1
\]

\[
t_2 = \frac{Y_i - X_i}{k} - \frac{G}{k^2} = \bar{t}_j + \bar{t}_i + \beta_2 + \varepsilon_2
\]

where \( \varepsilon_1 = (\bar{\varepsilon}_j + \bar{\varepsilon}_i - \bar{\varepsilon}_\cdot - \bar{\varepsilon}_1), \)

\[\varepsilon_2 = (\bar{\varepsilon}_i - \bar{\varepsilon}_j - \bar{\varepsilon}_\cdot - \bar{\varepsilon}_1)\]

and

\[
\beta_2 = \beta_x + \beta_y - \bar{\beta}_x - \bar{\beta}_y. \quad \text{Hence} \quad \frac{1}{W_1} \sim \sigma_1^2 = \frac{2(k-1)}{k^2} \sigma_0^2
\]

\[
\frac{1}{W_2} \sim \sigma_2^2 = \frac{2(k-1)}{k^2} \sigma_0^2 + \frac{2(k-1)}{k} \sigma_e^2 = \frac{2(k-1)}{k^2} \left[ \sigma_0^2 + k \sigma_b^2 \right].
\]

Therefore we can set \( \frac{1}{W_1} = \sigma_0^2 \) and \( \frac{1}{W_2} = \sigma_0^2 + k \sigma_b^2 \). If the experiment is large enough so that \( \sigma_0^2 \) can be replaced by \( V_0 \) and \( (\sigma_0^2 + k \sigma_0^2) \) by \((2V_b - V_0)\), the problem is solved. If there are fewer than 12 degrees of freedom to estimate \( \sigma_b^2 \), we do not recommend using this weighted analysis; instead, we advise using the unweighted analysis of section 9.7.2 (using no "inter-block information"). And with 12-16 degrees of freedom, the weights are not too stable. It is best to duplicate the experiment to obtain more information on \( \sigma_b^2 \) if weighting is to be used.
When computing adjusted treatment effects, some simplification can be made. We note that

\[ t_{ij} = \frac{x_{i+} + y_{i+}}{2} - \frac{x_{i+} + x_{i+} + y_{i+} + y_{i+}}{2k} + \frac{w_1(y_{i+} + y_{i+}) + w_2(x_{i+} + x_{i+})}{k(w_1 + w_2)} - m \]

\[ = \frac{T_{ij}}{2} + \frac{(w_1 - w_2)}{2k(w_1 + w_2)} (D_{ij} - D_{ii}) \]

Note that \( (D_{ij} - D_{ii}) \) is merely the negative of \( t_{ij} \) given on page 224. The only difference in \( t_{ij}' = t_{ij} + m \) is that \( \sigma \) is multiplied by the weighting factor \( \theta = (w_1 - w_2)/2k(w_1 + w_2) \) instead of \( 1/2k \). If \( \sigma \theta = 0 \), so that \( w_1 = w_2 \), no adjustment at all is made for blocks: \( t_{ij}' = T_{ij}/2 \). If \( \sigma \theta \) is small compared with \( \sigma^2 \), \( \theta \) approaches \( 1/2k \), the value used without recovery of inter-block information.

It can be shown that the variance of the difference between two adjusted treatment means, which appear in the same block, is

\[ \frac{1}{kw_1} \left[ (k - 1) + \frac{2w_1}{w_1 + w_2} \right] \]

And for varieties met in the same block, this variance is

\[ \frac{1}{kw_1} \left[ (k - 2) + \frac{4w_1}{w_1 + w_2} \right] \]

The average variance is

\[ \frac{1}{(k + 1)w_1} \left[ (k - 1) + \frac{4w_1}{w_1 + w_2} \right] = \frac{1}{w_1} \left[ 1 + \frac{4k\theta}{k + 1} \right] \]

Hence the efficiency of the simple lattice, relative to randomized complete blocks, is

\[ I = \frac{k + w_1/w_2}{(k - 1) + 4w_1/(w_1 + w_2)} \]
The reader is advised to read references (10) and (21) for more details on lattice designs.

If the simple lattice is duplicated several times (page 224) so that \(2d\) replications (2kd blocks) are used, it can be shown that the expected value of the "block effect" mean square with \(2(d - 1)(k - 1)\) degrees of freedom is \((\sigma_e^2 + k\sigma_b^2)\). The expected value of the other mean square with \(2(k - 1)\) degrees of freedom is, as above, \((\sigma_e^2 + \frac{3k}{2}\sigma_b^2)\). Hence if the two mean squares are pooled to obtain the adjusted blocks mean square with \(2d(k - 1)\) degrees of freedom, the expected value of this mean square is \((\sigma_e^2 + \frac{2d - 1}{2d} k\sigma_b^2)\). In this latter case,

\[
W_2 = \frac{2d - 1}{2dV_b - V_o} \quad \text{and} \quad \theta = \frac{d(V_b - V_o)}{2k \left[ 2V_b + (d - 1)V_o \right]}
\]

The variances of adjusted mean differences will all be divided by \(d\).

**Example 11.5.** In order to illustrate the computing procedure for recovery of inter-block information, we will use the example on page 223. Actually one should not contemplate any recovery with so few degrees of freedom for blocks,

\[
\begin{array}{cc}
D_i & D_j \\
\hline
-22 & 2 \\
-21 & -16 \\
3 & -32 \\
46 & -46 \\
\end{array}
\]

SSB(adj.) = \(\sum (D_i^2 + D_j^2) - \frac{(-46)^2}{9} = 134\)

As a check, SSB (unadj.) + SST (adj.) should equal SSB (adj.) + SST (unadj). We note that

\(68 + 593 = 134 + 527 = 661\).

It is advisable to make this check on the computing. Also you can test for treatment differences using SST (adj.). The analysis of variance is:
<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>E(MS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replications</td>
<td>1</td>
<td>118</td>
<td></td>
</tr>
<tr>
<td>Treatments (adj.)</td>
<td>8</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>Blocks (adj.)</td>
<td>4</td>
<td>34 = V_b</td>
<td>$\sigma^2_b + 1.5 \sigma^2_e$</td>
</tr>
<tr>
<td>Error</td>
<td>4</td>
<td>47 = V_e</td>
<td>$\sigma^2_e$</td>
</tr>
</tbody>
</table>

The estimate of $\sigma^2_b$ would be negative so we assume $\sigma^2_b = 0$. Hence, no adjustments would be made even if there were sufficient degrees of freedom to estimate the weights. This raises a rather critical issue regarding incomplete blocks designs; namely, that we do not take the loss in efficiency which we really suffered by using such a design with a negative estimate of $\sigma^2_b$. Instead we use $t_{ij}' = T_{ij}/2$, with the variance of a mean equal to $47/2 = 24$.

If $V_b > V_e$ so we could obtain a positive estimate of $\sigma^2_b$ and if there were sufficient degrees of freedom to warrant weighting, we would compute $\bar{\bar{\theta}} = D_{ij} - D_{i.}$ (see page 224). Then $\theta = \frac{W_1 - W_2}{6(W_1 - W_2)}$, where $W_1 = \frac{1}{V_e}$ and $W_2 = \frac{1}{2V_b - V_e}$. Here we assume that $\theta = 0$. An adjusted treatment mean is

$$t_{ij} = \frac{T_{ij}}{2} + \theta(D_{ij} - D_{i.}) = \frac{T_{ij}}{2} - \theta \bar{\bar{\theta}}_{ij}.$$

The estimated average variance of the difference between two adjusted means is

$$\frac{V_e}{4} \left[ 2 + \frac{4W_1}{W_1 + W_2} \right].$$
11.3.4. Other Lattice Designs. Reference (21) contains the theory and computing details for a triple lattice. One of the present authors presented the method of analysis for a duplication of any orthogonal square lattice design. If there are \( r \) replications in the basic design (rd total replications) and if the two parts of the adjusted blocks sum of squares are pooled with \( rd(k - 1) \) degrees of freedom,

\[
W_2 = \frac{rd - 1}{rdV_p - V_e} \quad \text{and} \quad \theta = \frac{W_1 - W_2}{rk[(r - 1)W_1 + W_2]}. 
\]

The average variance of the difference between two adjusted means is

\[
\frac{2V_e}{dr} \left[ 1 + \frac{r^2k}{k + 1} \theta \right].
\]

11.3.5. Balanced Incomplete Blocks Designs. Yates (22) presented the basic ideas of recovery for balanced incomplete blocks designs. The inter-block estimate of \( \overline{\tau}_i \) (see section 9.6) is obtained from \( T_{bi} = \sum_j n_{ij} \beta_j \).

\[
T_{bi} = \frac{k\theta}{p} = (r - \lambda) \overline{\tau}_i + k \sum_j n_{ij} \beta_j - \frac{k^2}{p} \sum \beta_j + \sum_i \sum_j n_{ij} n_{ij} \epsilon_{ij} - \frac{k}{p} \sum \sum n_{ij} \epsilon_{ij} \]

\[
\overline{\tau}_i = \frac{T_{bi} - \frac{k\theta}{p}}{r - \lambda} \equiv t_{bi}
\]

\[
\sigma^2(t_{bi}) = \frac{rk(p - k)}{p} \left( \frac{k \sigma^2}{\theta} + \frac{\sigma^2}{\theta} \right) = \frac{k(p - 1)}{p(r - \lambda)} \frac{1}{W_2} \sim \frac{k}{(r - \lambda)W_2}
\]

Since \( \sigma^2(t_1) = \frac{(k - 1)}{kE_f} \frac{1}{rE_fW_1} = \frac{p - 1}{p} \frac{1}{rE_fW_1} \sim \frac{1}{rE_fW_1} \), where \( t_1 = \frac{kT_1 - T_{bi}}{rkE_f} \), we can set up the following weighted estimate of the adjusted treatment mean:

\[ t_i' = \frac{kT_i - T_{bi}}{rE_F W_1} + \frac{T_{bi}}{(r - \lambda)k} - \frac{(r - \lambda)W_2}{p(r - \lambda)k} - \frac{kG}{rp} \]

\[ = \frac{T_i}{r} + \theta D_i, \]

where \( \theta = \frac{W_1 - W_2}{r[p(k - 1)W_1 + (p - k)W_2]} \) and \( D_i = \left[ (p - k)T_i - (p - 1)T_{bi} + (k - 1)G \right]. \)

Certain identities are useful in making the above simplification:

\( (r - \lambda)(p - 1) = r(p - k) \) and \( kE_F(p - 1) = p(k - 1) \).

The variance of the difference between two adjusted means, using inter-block information, is

\[ \frac{2k(p - 1)}{r[p(k - 1)W_1 + (p - k)W_2]} . \]

The computations needed to estimate \( W_1 \) and \( W_2 \) are the following:

(i) SSB (unadj.) in the usual manner

\[ \frac{\sum_i T_{bi}^2}{k(r - \lambda)} - \frac{(\sum T_{bi})^2}{pk(r - \lambda)} = SST_b \]

(ii) \[ \sum_i D_i^2 = SSD \]

(iii) \[ \frac{\sum_i D_i^2}{N(p - k)(k - 1)} = SSD \]

(iv) SSB (adj.) = SSD + SSB (unadj.) - SST_b, with \( (q - 1) \) degrees of freedom.

\[ V_b = MSB (adj.) \]

(v) \[ SSE = Sy^2 - SSB (adj.) - SST (unadj.), with N - p - q + 1 degrees of freedom. \]
(vi) Since $E(V_b) = \sigma_e^2 + \frac{N - p}{q - 1} \sigma_b^2$,

$$W_2 = \frac{N - p}{k(q - 1)V_b - (p - k)V_e}$$

(vii) Compute $r\theta$ and then $t'_1 = (T_1 + r\theta D_1)/r$.

In general it is also advisable to compute SST (adj.) as a check on the computation and to make a test of the differences among the treatment means.

If groups of blocks form complete replications, the sum of squares for replications with $(c - 1)$ degrees of freedom is removed from SSB (adj.) and the expected value for $V_b$ (with $q - c$ degrees of freedom) is

$$E(V_b) = \sigma_e^2 + \frac{N - p - k(c - 1)}{q - c} \sigma_b^2.$$ 

Hence

$$W_2 = \frac{N - p - k(c - 1)}{k(q - c)V_b - (p - k)V_e}.$$ 

If $c = r$, $E(V_b) = \sigma_e^2 + \frac{k(r - 1)}{r} \sigma_b^2$ and $W_2 = \frac{r - 1}{rV_b - V_e}$.

This is the result for balanced lattices with $d = 1$.

Example 11.6. Suppose we estimate $\sigma_b^2$ from the balanced lattice experiment of exercise 9.6.1. The values of $D_1 = 6T_1 - 8T_{b1} + 2G$ are 242, -6, -276, 118, 74, -28, -80, -76, and 32.

$$SST_b = \frac{(152)^2 + \cdots + (188)^2}{9} - \frac{(1512)^2}{81} = 127$$

$$SSD = \frac{(242)^2 + \cdots + (32)^2}{432} = 389$$

$$SSB (adj.) = 389 \left[ \text{since } SST_b = SSB \text{ (unadj.)} - SS \text{ (Rops.)} \right].$$
\[ V_b = \frac{389}{8} = 48.63; \quad V_e = 28.57 \]

\[ W_2 = \frac{3}{4V_b - V_e} = \frac{3}{4(48.63) - 28.57} = 0.01808 \]

\[ W_1 = \frac{1}{V_e} = 0.03500 \]

\[ \epsilon = \frac{W_1 - W_2}{4(18W_1 + 6W_2)} = \frac{V_b - V_e}{72V_b} = \frac{20.06}{3501} = 0.00573 \]

The adjusted treatment means, \( t_i^* = \frac{T_i}{r} + 6D_i \) are 20.13, 11.72, 2.92, 22.93, 14.17, 11.34, 13.54, 7.06, 22.18. The variance and standard error of the difference between two adjusted means are

\[ s^2(d) = \frac{48}{4(18W_1 + 6W_2)} = \frac{2}{3W_1 + W_2} = 16.25 \]

\[ s(d) = 4.03 \]

Using recovery of inter-block information, the relative efficiency of this design compared to randomized complete blocks is \( I = 17.64/16.25 = 1.086 \). Hence the recovery resulted in a gain of 16 per cent over no recovery. However with only 8 degrees of freedom to estimate \( \sigma^2_b \), we believe that this is a rather fictitious gain since the weights \( W_1 \) and \( W_2 \) themselves have considerable error.

11.3.6 Covariance. Cochran (23) considered the following model for a sampling design with \( p \) plots and \( r \) samples per plot:

\[ y_{ij} = \mu + \tau_i^* + \beta x_{ij} + \epsilon_{ij}, \]

where the \( \tau_i^* \)'s and \( \epsilon_{ij} \)'s are assumed NID with zero means and respective variances \( \sigma^2_\tau \) and \( \sigma^2_\epsilon \), and \( \mu, \beta \) and the \( x \)'s are assumed fixed. The maximum likelihood estimate of \( \beta \) is a weighted mean of the \( b \)'s derived from the
plots \((b_t)\) and error \((b_e)\) lines of the analysis of covariance (see section 10.2.5 with treatments corresponding to plots).

If information about \(\beta\) in the plots line is disregarded, the procedure is as follows:

(i) \(s^2\) is an unbiased estimate of \(\sigma^2\) and \(SST^*\) is distributed as \(\chi^2_{\sigma^2}(r-1)\).

(ii) \(SST^* = r \sum_1^r (\bar{Y}_{i*} - \bar{Y} - b_t \bar{x}_{i*})^2 + (b_t - b_e)^2 \frac{T_{XX}E_{XX}}{T_{XX} + E_{XX}}\), where

\[\bar{Y}_{i*} = \mu + \tau^*_i + \beta \bar{x}_{i*} + \bar{e}_{i*}\]. The first term is distributed as

\((\sigma^2 + r \sigma^2_t) \chi^2_{p-2}.\) Since

\[b_t = \beta + r \sum_1^r (\tau^*_i + \bar{e}_{i*}) \bar{x}_{i*}, \quad b_e = \beta + \sum \sum \bar{e}_{i*} (x_{i1} - \bar{x}_{i*}),\]

the second term is distributed as \((\sigma^2 + \lambda r \sigma^2_t) \chi^2_p\), where

\[\lambda = E_{XX}/(T_{XX} + E_{XX}).\]

(iii) All three \(\chi^2\) are independent.

(iv) \(\hat{\sigma}_r^2 = \frac{(MST^* - s^2)}{r} \frac{(p-1)}{(p-2+\lambda)}\).

The estimate of \(\sigma_r^2\) in (iv) involves some loss of information because plot information about \(\beta\) is ignored, and the single degree of freedom in \(SST^*\) is presumably given too much weight.

If the estimate of \(\beta\) from the error line is used, parallel results can be obtained for a two-way classification, e.g. randomized blocks.

**Exercise 11.3.1.** Prove that the expected value of \(s^2\) in 9.3.3 is actually

\[\sigma^2_o + \sigma^2_{tb}\] if there is an interaction.
Exercise 11.3.2. (a) Show how the coefficients of the variance components in example 11.3 were obtained.

(b) What is the variance of the average amount of honey per colony per year in a given row if there are $g$ colonies and $r$ years? What is the expected value of this average amount of honey in terms of the parameters given in the example?

(c) What is the variance of the difference between the averages for any two rows?

Exercise 11.3.3. (a) In example 11.4, compute the estimates of the variance components, and also determine approximate 90 per cent confidence limits for these estimates, using one of the last three methods in section 11.2.1.

(b) Make the necessary tests of significance for both fixed and random effects.

(c) Determine 15 treatment-year means and the 5 treatment and 3 year means. What are the standard errors of the differences for the 5 types of differences mentioned in the text? Illustrate each difference by a specific example from the above 23 means.

(d) How would the split-plot analysis be changed if the ($\alpha \beta$) interaction were included in the model and were assumed to be a random effect?

Exercise 11.3.4. (a) In exercise 9.5.1 (page 201), what changes should be made in the analysis if the four blocks and five varieties constituted representative samples from large universes of blocks and varieties?

(b) What are the expectations of the mean squares under this assumption?

(c) Determine estimates of the random variance components and make the necessary tests of significance.

(d) Make a test of spacing differences and determine the standard error of the difference between two spacing means if $p$ varieties and $r$ blocks were used.

Exercise 11.3.5. A field experiment was conducted to determine an acceptable lower limit on the size of similar future experiments. The analysis of variance was:
<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean Squares</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocks</td>
<td>2</td>
<td>4,399</td>
</tr>
<tr>
<td>Treatments</td>
<td>4</td>
<td>4,429</td>
</tr>
<tr>
<td>T x B</td>
<td>8</td>
<td>866</td>
</tr>
<tr>
<td>Samples in plots</td>
<td>15</td>
<td>239</td>
</tr>
<tr>
<td>Determinations in Samples</td>
<td>30</td>
<td>7</td>
</tr>
</tbody>
</table>

(a) Assuming everything random, except treatments, set up the expectations of the mean squares and determine estimates of the random components.

(b) What is the standard error of the difference between two treatment means with \( r \) blocks, \( k \) samples per plot, and \( d \) determinations per sample? Which of these three sampling plans (which cost the same) would be favored: \( r = 6, k = 2, d = 1; r = 3, k = 5, d = 1; r = 4, k = 2, d = 2 \)?

**Exercise 11.3.6.** Suppose data were available on average sales price per pound of tobacco in North Carolina for a period of 4 years at each of 38 markets. These markets are divided into 4 geographical areas: 8 in area A, 9 in B, 14 in C and 7 in D.

(a) Set up an analysis of variance to reflect the sources of variation in the marketing picture, assuming that areas are fixed and the remaining components are random.

(b) Determine the expectations of the mean squares.

(c) What is the variance of a state mean over all 4 areas with \( m \) markets per area, and \( y \) years?

**Exercise 11.3.7.** Johnson and Tsao (24) conducted a psychological experiment to determine the difference limen (D.L.) of subjects for weights increasing at constant rates. Two classes of people were chosen as to sight—normal (A) and congenitally blind (B). Two males and two females were selected to represent each class, giving a total of 8 people. Then the average of five D.L. values was determined for each subject for each of 28 rate-weight combinations, 7 weights and 4 rates. The entire experiment was repeated at a later date. Hence there was a total of \( 28 \times 8 \times 2 = 448 \) average D.L.'s. The order of presentation of the 28 combinations was randomized.
for each subject at each date.

(a) Set up the analysis of variance as to sources of variation, degrees of freedom, and expectations of the mean squares.

(b) How would the analysis be changed if each of the five D.L. values were recorded instead of a single average being reported?

(c) A portion of the data (for three weights and 2 rates) is reproduced below for computational purposes:

<table>
<thead>
<tr>
<th>Weight Rate</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>M</td>
<td>B</td>
<td>F</td>
</tr>
<tr>
<td>Sight</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>Date</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Individual</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>4.5</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>2</td>
<td>14.0</td>
<td>12.2</td>
<td>8.1</td>
</tr>
<tr>
<td>1</td>
<td>3.1</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>2</td>
<td>13.4</td>
<td>13.8</td>
<td>12.8</td>
</tr>
<tr>
<td>1</td>
<td>24.2</td>
<td>25.3</td>
<td>24.1</td>
</tr>
<tr>
<td>2</td>
<td>19.3</td>
<td>24.1</td>
<td>19.6</td>
</tr>
<tr>
<td>1</td>
<td>18.5</td>
<td>22.3</td>
<td>11.4</td>
</tr>
<tr>
<td>2</td>
<td>3.1</td>
<td>3.9</td>
<td>3.6</td>
</tr>
<tr>
<td>1</td>
<td>11.2</td>
<td>8.8</td>
<td>8.1</td>
</tr>
<tr>
<td>2</td>
<td>3.9</td>
<td>5.1</td>
<td>3.7</td>
</tr>
<tr>
<td>1</td>
<td>9.6</td>
<td>7.3</td>
<td>6.5</td>
</tr>
<tr>
<td>2</td>
<td>9.0</td>
<td>6.4</td>
<td>6.9</td>
</tr>
<tr>
<td>1</td>
<td>8.6</td>
<td>13.9</td>
<td>12.4</td>
</tr>
<tr>
<td>2</td>
<td>8.6</td>
<td>14.5</td>
<td>5.3</td>
</tr>
</tbody>
</table>

Exercise 11.3.8. In section 11.3.3, prove that $v_o$ as therein defined is an unbiased estimate of $\sum \frac{2}{e}$. In other words, show that

$$SSE = Sa^2 + Sy^2 - SST (unadj.) - SSB (adj.) .$$

Exercise 11.3.9. Prove that, if we have two unbiased independent estimates, $t_1$ and $t_2$, of parameter $\tau$ with respective variances $\frac{1}{W_1}$ and $\frac{1}{W_2}$, the combined linear unbiased estimate with lowest variance will be

$$t = \frac{W_1 t_1 + W_2 t_2}{W_1 + W_2}$$

Hint: remember that $E(t) = \tau$. 

-298-
Exercise 11.3.10. On page 287, show that the results for $\sigma_1^2$ and $\sigma_2^2$ are correct. Why is $(2V_b - V_o)$ an unbiased estimate of $(\sigma_0^2 + k \sigma_b^2)$?

Exercise 11.3.11. Derive the formulas given on page 288 for the variances of the difference between two adjusted means, using a simple lattice design. Also derive the formula for relative efficiency, $I$.

Exercise 11.3.12. (a) Analyze the data in exercise 9.7.1, using recovery of inter-block information.

(b) What is the relative efficiency of this design compared to a randomized complete blocks design?

Exercise 11.3.13. (a) In section 11.3.5, derive the formula for $\sigma^2(t_{b_1})$, and show that $t_{i_1} = T_{ij} / r + \sigma_{b_1}$, as presented there.

(b) Also derive the variance of the difference between two adjusted means.

Exercise 11.3.14. Analyze the data in exercises 9.6.2 and 9.6.3 by the methods of section 11.3.5.

Exercise 11.3.15. (a) Prove Cochran's results for the use of variance components in a covariance analysis and apply them to exercise 10.5.

(b) Derive the same results for a randomized complete blocks experiment.

11.4. The Use of Components of Variance in Regression Problems.

Variance components are also used to evaluate methods of estimating regression coefficients. In Chapter 8, we assumed that for every $X$ there was only one value of $Y$. The experimenter might want to obtain several values of $Y$ for each value of $X$ in order to determine the sampling or observational error. The model for $X$ independent variates with $p$ samples of $Y$ for each set of $\sum_i X_i$ values is

$$Y_{jk} = \mu + \sum_{i=1}^r \beta_i x_{ij} + \epsilon_j + \sigma_{jk}, \quad (j = 1, 2, \ldots, n),$$

$$\sigma_{jk} = \sigma^2, \quad (k = 1, 2, \ldots, p)$$
where $\varepsilon$ and $\eta$ are NID with zero means and respective variances $\sigma^2_\varepsilon$ and $\sigma^2_\eta$. $\sigma^2_\eta$ measures the failure of the regression line to go through the average values of $Y(\bar{Y}_j)$, while $\sigma^2_d$ measures the fluctuation of individual values $Y_{jk}$ about their means $\bar{Y}_j$.

The analysis of variance of the residuals will have the following appearance:

<table>
<thead>
<tr>
<th>Error</th>
<th>D.F.</th>
<th>Mean Square</th>
<th>EMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviation from regression</td>
<td>$n - r - 1$</td>
<td>$V_e$</td>
<td>$\sigma^2 = \sigma^2_d + p \sigma^2_\varepsilon$</td>
</tr>
<tr>
<td>Observations</td>
<td>$n(p - 1)$</td>
<td>$V_d$</td>
<td>$\sigma^2_d$</td>
</tr>
</tbody>
</table>

In this case the experimenter can determine if he should add more sets of $X$'s or collect more values of $Y$ for each set, basing his conclusions on the relative size of $\sigma^2_d$ and $\sigma^2_\varepsilon$ as compared to the relative costs of obtaining more sets or more observations per set.

Very few examples are available of planned experiments with the same number of observations per set (other than one observation). In many cases, there will be experiments with several values of $Y$ for a given $X_1$ but the values of $X_2$, $X_3$, ..., vary as well as $Y$ (see page 116). Hence we are led to use the one-error regression equation. Another difficulty often arises; namely, that the observation error tends to increase with increasing $\bar{Y}$. This difficulty is not avoided when we use only one observation per set, but it is usually neglected (see section 8.7). If several observations were taken for each set of $X$ values, the experimenter would have some information as to the uniformity of his variation as $\bar{Y}$ increased.

The factorial design is an example of several observations for each set of $X$'s; however, the $X$ values are often qualitative rather than quantitative (different varieties, cultivation methods, teaching methods, etc.). When a factorial is used with different levels of the factors, the principle of multiple regression with several observations for each set of $X$ values is being applied. However, there
are seldom enough levels of the various factors to estimate the deviation error. (Exercises 9.5.2 and 9.5.4 are examples). The article referred to in Exercise 9.5.7, was based on estimating a quadratic regression of yield on planting date of cotton with 10 equally-spaced planting dates. In this case \( \sigma_d^2 \leq 0.3606 \) and 
\[ \sigma_d^2 + p \sigma^2_\epsilon \leq 0.3357, \] a value less than the estimate of \( \sigma_d^2 \). This is an all too common result - the estimate of \( \sigma_\epsilon^2 \) is negative. This indicates a need for a more thorough investigation for many regression problems of the accuracy of the assumptions of homogeneous variance from point to point and independence of the true residuals. The authors have not been able to find many simple exercises for this section; if any are available, we would appreciate receiving them.

**Example 11.7.** A survey was conducted in Eastern North Carolina in 1949 to estimate the relationship between per cent dry weight (Y) of Irish potatoes (Cobbler variety) and \( X = 200 \) (specific gravity - 1.045). The values of \( X \) were -1, 0, 1, • • •, 11 with 8 samples and 2 determinations per sample for many samples. If there had been potatoes for each of the 13 \( X \)-classes for each sample and determination, there would have been 104 samples and 208 determinations. Unfortunately, there were few samples for \( X = -1 \) and \( X = 11 \) and many blanks in other places, so that only 127 determinations were obtained. The data were as follows:

<table>
<thead>
<tr>
<th>Sample</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6*</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14.34</td>
<td>14.73</td>
<td>15.30</td>
<td>17.46</td>
<td>18.82</td>
<td>19.48</td>
<td>20.40</td>
<td>20.67</td>
<td>21.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.66</td>
<td>14.87</td>
<td>16.79</td>
<td>17.23</td>
<td>17.68</td>
<td>19.16</td>
<td>19.90</td>
<td>21.27</td>
<td>22.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.12</td>
<td>15.04</td>
<td>16.50</td>
<td>16.75</td>
<td>17.90</td>
<td>18.91</td>
<td>20.35</td>
<td>20.72</td>
<td>22.28</td>
<td>23.45</td>
<td>24.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.82</td>
<td>15.88</td>
<td>16.43</td>
<td>17.08</td>
<td>18.39</td>
<td>19.32</td>
<td>20.89</td>
<td>22.28</td>
<td>23.31</td>
<td>24.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.58</td>
<td>15.48</td>
<td>15.80</td>
<td>17.38</td>
<td>17.86</td>
<td>18.81</td>
<td>19.68</td>
<td>20.78</td>
<td>21.83</td>
<td>23.50</td>
<td>24.60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.83</td>
<td>15.04</td>
<td>16.17</td>
<td>17.63</td>
<td>17.90</td>
<td>19.15</td>
<td>20.11</td>
<td>20.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.92</td>
<td>15.36</td>
<td>16.10</td>
<td>16.68</td>
<td>17.88</td>
<td>20.18</td>
<td>21.14</td>
<td>21.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15.30</td>
<td>15.60</td>
<td>16.63</td>
<td>17.98</td>
<td></td>
<td>19.38</td>
<td>20.56</td>
<td>22.78</td>
<td>24.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.09</td>
<td>14.10</td>
<td>15.50</td>
<td>16.48</td>
<td>17.63</td>
<td>19.10</td>
<td>20.34</td>
<td>21.71</td>
<td>22.73</td>
<td>23.60</td>
<td>25.56</td>
<td></td>
</tr>
<tr>
<td>8**</td>
<td>13.86</td>
<td>14.63</td>
<td>15.14</td>
<td>16.49</td>
<td>17.10</td>
<td>18.51</td>
<td>19.18</td>
<td>20.26</td>
<td>21.44</td>
<td>22.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.61</td>
<td>14.63</td>
<td>14.90</td>
<td>16.27</td>
<td>17.20</td>
<td>17.97</td>
<td>19.08</td>
<td>20.80</td>
<td>21.64</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*3 determinations on the second sample; **There was one determination = 13.05 for \( X = -1 \).
The analysis for the third sample is as follows:

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>( \sigma^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>1</td>
<td>253.593</td>
<td>( \sigma_d^2 + 2 \sigma_e^2 )</td>
</tr>
<tr>
<td>Deviations</td>
<td>9</td>
<td>0.1086</td>
<td>( \sigma_e^2 )</td>
</tr>
<tr>
<td>Determinations</td>
<td>1.11</td>
<td>0.1269</td>
<td>( \sigma_d^2 )</td>
</tr>
</tbody>
</table>

Again we note a negative estimate of \( \sigma_e^2 \). For the five samples with duplicate determinations, two had negative estimates of \( \sigma_b^2 \), two had positive estimates and one was almost zero (slightly positive).

An over-all regression analysis for all eight samples gave \( \sigma_d^2 = 0.1362 \) and \( \sigma_d^2 + \lambda \sigma_e^2 = 0.1804 \). The value of \( \lambda \) is somewhere between 1. and 2. A rough approximation is the number of determinations divided by the number of samples = 127/79 = 1.6.

Exercise. (a) Determine the value of \( \lambda \) by use of the regression model.
(b) Estimate \( \sigma_d^2 \) for the other four samples with duplicate determinations.
(c) Show how the over-all estimates of \( \sigma_d^2 \) and \( \sigma_d^2 + \lambda \sigma_e^2 \) were obtained.
(d) How would you set up a computing procedure to test the over-all regression and the deviations of sample regressions from the over-all regression?

REFERENCES CITED


OTHER REFERENCES

General


Experimental


Sample Surveys


